Chapter 1
Introduction

1.1 Image Registration

Image Registration in Computer Vision: Human beings have a tendency to gather information from their surroundings. They observe, evaluate and accordingly define their actions. Similarly, Computer vision aims in interpreting human vision through images captured using electronic devices. Applications of computer vision involves extraction, formulation and use of image information in artificial intelligence, remote sensing, medical image analysis, augmented reality, surveillance system and many other applications. In these listed applications of computer vision, extracting useful information in form of scene shape, structured information, color content etc. from images becomes important for performing task like image registration, object tracking etc. [Azuma et al. 2001, Sharma and Goyal 2013].

Definition and purpose of Image Registration: Image Registration by definition could be explained as the alignment of same scene in different images. These images may capture same scene in varied conditions like change of viewpoint, illumination, scale etc. Therefore, the process of aligning the images and compensating for the differences caused by different image acquisition conditions is called image registration [Wyawahare et al. 2009, Wang et al. 2017]. Figure 1.1 represents the basic concept of image registration procedure where one-to-one mapping between the coordinates in one image space is mapped to those in another i.e. points belonging to two image spaces that correspond to the same scene point are mapped to each other (example, in Figure 1.1, point (x_0, y_0) in Image (a) maps to point (x_1, y_1) in Image(b)).

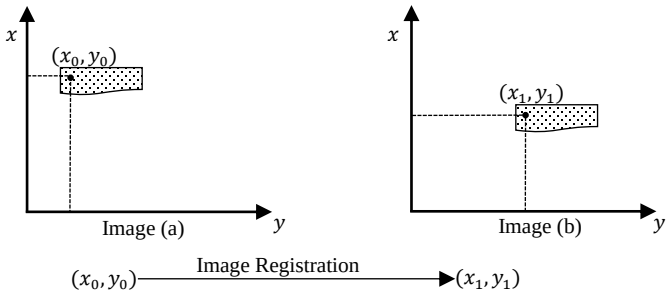


Fig. 1.1. Image Registration

Image Registration takes upon applications related to various computer vision and pattern recognitions tasks like image segmentation, shape reconstruction, motion tracking, object tracking, medical image analysis etc.

Classification of Image Registration: Classification of Image Registration, as described by a survey [Wyawahare et al. 2009], could be based on features like: 1) dimensionality involving Two Dimensional (2D)-2D, 2D-Three Dimensional (3D) or even 3D-3D image registration, 2) nature of registration i.e. source of features determining the availability of desired features within the image (intrinsic features) or explicitly introduced external objects in an image for processing (extrinsic features), 3) nature of transformation describing the rigid, affine, projective or nonlinear transformation, 4) domain of transformation where formulation takes into consideration the whole image (global) or chooses a part of the image (local) for registration and 5) subject of registration determining intrasubject: two images containing the same subject or intersubject: two images containing different subjects.

Applications of Image Registration: Image Registration applications could be broadly classified in categories as: 1) remote sensing, involving tasks like multispectral classification, environmental monitoring, image mosaicing, weather forecasting, Geographic Information Systems (GIS) etc., 2) medicine, where computer tomography and Nuclear Magnetic Resonance (NMR) spectroscopy data could be used to determine complete information about the patient, tumor growth could be monitored, patient's data could be compared with anatomical atlases etc., 3) cartography, performing map creation and updating and 4) computer vision, involving tasks like target localization, automatic quality control, image segmentation, object tracking etc.

1.2 Augmented Reality

Concept: Augmented Reality (AR) aims in enhancing the visualized view of user's environment by combing real and virtual information together in the view of real environment, where the virtual information comprises of visualized objects, graphics or sound effects. AR superimposes real world objects with virtual objects or allows them to be rendered in the same scale space, giving an illusion to the user as if the virtual rendering is a part of the real environment, i.e., view of real scene is augmented with additional virtual information.

Caudell and Mizell in 1992 [Caudell and Mizell 1992] introduced an application of AR for pilot training and since then, AR has been used for variety of other fields. One such example is briefly represented in Figure 1.2, where AR is used for generating instructions to perform a machine repairing task. To obtain such functionality, an image database is prepared, where each image in the database serves

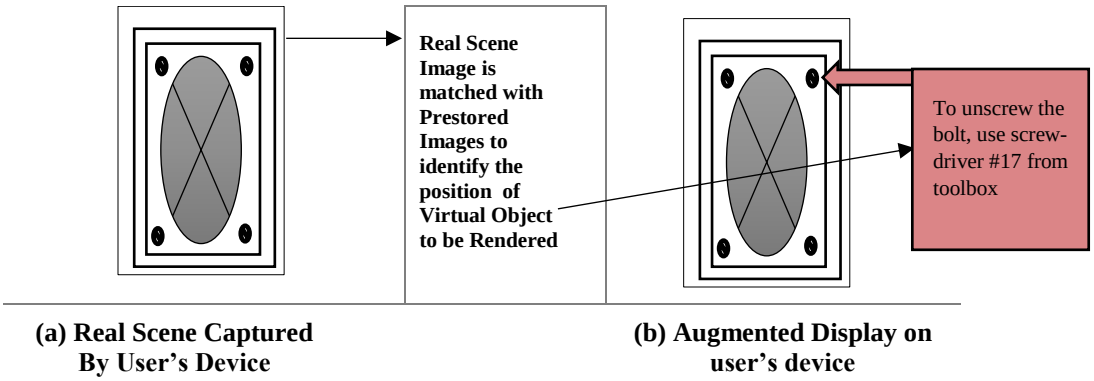


Fig. 1.2. An Example of Machine Repairing Using AR

as a reference image to the real time captured image or video frame. Now the captured image is matched with each image in the database for evaluating the correct position of the virtual information (here in example, the position of a bolt to unscrew to further open the machine) and the kind of virtual information (here in example, a message to select an appropriate screw-driver) that is to be augmented in the view of real scene. Figure 1(b) shows the augmented image, where the added information appears with the real scene when viewed on user's device. Google Glass [Miller 2013] is also an AR example which allows the user to access mobile functionalities through voice commands and functions as a hands-free smartphone.

Applications: Initially, implications of AR mainly focused on gaming applications, but the uses of AR over the last one decade became much broader. Applications of AR can be widely listed including areas like visual medical surgery, military use etc. [Clarke et al. 2003, Lin et al. 2006]. Also, possible commercial and educational applications [Fleck et al. 2013, Ibáñez et al. 2013, Sergey et al. 2015, Bai et al. 2015, Akçayır et al. 2016] of AR are unlimited. Gupta and Rohil [Gupta and Rohil 2017b] provides an overview of few possible applications of AR in different fields of education and training. For example: understanding concepts like mutation of genetic materials, evolution of living beings, functionality of human body, astronomy and holographic universe etc. could be made more fun and easy by using AR to provide students a way to visualize the written content by adding corresponding graphics. Also, some students find it difficult to comprehend and learn few basic concepts of physics such as projectile motion where remembering a number of formulae for solving a particular problem becomes difficult. Therefore, the learning process could be made more effective by adding real-time examples with audio-video effects in their actual environment with some pop-up windows that could suggest possible solutions for a particular problem. This would make the students learn quickly and have a firmer grasp on concepts.

Image Registration for Augmented Reality: AR integrates digital information with the user's environment in real time. In simple words, AR takes an existing picture and blends new information into

it [Clarke et al. 2003]. It is a reconstructed view of a physical and real-world environment where additional information is blended using computer-generated details such as sound, video, graphics, image or spatial data. Various technologies such as stationary monitors or Head Mounted Display (HMD) etc. are used for visualizing the generated graphical overlay. Correct positioning of real objects and virtual data is a fundamental feature of an AR system. In AR systems, ideally the virtual and real objects must appear to co-exist in the same space and merge seamlessly. Therefore, even when user moves his/her head (in case of AR systems using HMD) the graphics should remain aligned to the real objects. This makes accurate registration of real and virtual worlds, one of the key requirements in designing an AR system. The process of registration can be simplified in three steps [Lin et al. 2006]:

Step 1. Positioning; to estimate observer's location and orientation so that virtual data could be added to the real environment correctly.

Step 2. Rendering; to obtain 2D projected image from 3D model which corresponds to a real image of the 3D model seen by the observer.

Step 3. Merging; is an image processing procedure to merge the virtual objects and real environment in order to make virtual objects look like real part of the scene.

Besides achieving the objective of seamless blending of virtual objects in the view of real environment, another key objective of an AR system could be defined as the accurate and precise alignment of virtual objects in the view of real environment in real time. AR system when built for consumer application, for example, in cases when AR is to be developed on a personal computer or on a mobile device, achieving real time performance becomes a bigger issue.

Design of an AR system comprises of three main steps where:

- The first step aims in correct pose evaluation of virtual objects that are to be rendered in the real scene. This step comprises of image registration where orientation and location of virtual objects are approximated with respect to the real environment using direction and location finding equipments such as Global Positioning System (GPS), Gyrometer etc. for finding the six-degree-of-freedom (6DOF) of a user or using Computer Vision methods or by performing hybrid registration that combines the ascendancy of direction finding equipment and Computer Vision methods.

  All these registration procedures have their own advantages and disadvantages in terms of speed, accuracy, computational complexity etc. For example: registration by GPS and Gyro-meter has faster speed but much lower accuracy as compared to registration based on Computer Vision methods. One such example of Computer Vision methods used for performing image registration is the Template matching method (i.e. searching for the most similar image pattern in the image with a template image obtained from prearranged information) which is considered to be a profitable approach to solve

problems of locating and recognizing an object in an image [Lin et al. 2009]. The basis of template matching methods can be classified in two categories:

1) Correlation based template matching: works with raw or semi-raw intensity images using a measure of similarity between template and image pixels. Correlation is the basic statistical approach to registration. It is often used for template matching or pattern recognition in which the location and orientation of a template or pattern is found in an image. By itself, correlation is not a registration method, it is a similarity measure or match metric, i.e., it gives a measure of the degree of similarity between an image and a template. However, there are several registration methods for which it is the primary tool and these methods are generally useful for images which are misaligned by small rigid transformations (i.e. shape of the object does not change during or after transformation) [Martedi et al. 2013]. Moreover, a number of other techniques have been developed by researchers to measure similarities between an input image and the reference template. Generally, similarity measure based accuracy evaluation depends upon the type of method or algorithm selected, kind of problem which is to be solved and the type of template and application it is to be used. In addition, sensitivity to the noise in the image, changing image luminance and computational complexity are some of the added factors considered while selecting a particular similarity measure.

2) Geometrical Image Descriptors based template matching: works with prior known features of the target like characteristics, lines or edges.

Many Computer Vision registration methods require placing of markers in the scene beforehand (marker based AR) and these methods work comparatively well for indoor environments as marker-based approaches are very sensitive to outdoor lightning and usually fails if the environment is not sufficiently illuminated [Clarke et al. 2003]. Also these approaches become unconvincing and less practical when markers are hidden from view by other moving or stationary objects. Therefore, markerless registration in an AR system has been recognized as a cogent research area [Genc et al. 2002, Lin et al. 2009] to obtain accurate and robust image registration results in an indoor as well as outdoor environment. AR systems based on computer vision techniques make use of certain algorithms to calculate pose and orientation of camera for accurate virtual overlay. This approach doesn't make use of any sensors and provide a better user experience using natural features from the scene or position and orientation of man-made markers added to the scene for estimating the pose accuracy. Computer vision based AR systems are classified in two categories: Marker based and Markerless AR systems.

Marker Based Augmented Reality: Marker based AR systems use markers (example: as described in ARToolkit [Miller 2013], ARTag [Fiala 2005] etc.) as a recognizable sign in the scene in form of a distinctive predefined pattern or a high contrast image preferably having significant number of edges

and very less repetitive patterns, to process the pose evaluation of virtual objects that are to be rendered in the view of real scene. Image or any form of pattern used as a marker is known best if it incorporates high distinctive properties that differentiates the marker from the real environment so that extracting and tracking the marker in subsequent image frames becomes less expensive in terms of computational complexity. Marker based AR makes the pose evaluation process easy but are not practical & convenient in all scenarios.

Markerless Augmented Reality: Markerless AR extracts natural features from the real scene and use them to evaluate the pose and orientation of virtual objects that are to be integrated in the view of real environment. Extraction of stable image features form a basis of computation in various computer vision tasks like robot navigation, image retrieval, 3D scene construction, building an AR system etc. These image features are expected to be highly stable, invariant to affine transformations and other imaging conditions like scale, blur and illumination change. Such features are often regions with some distinguished properties from their neighboring pixels determining them as regions of interest [Lin et al. 2009]. These detected regions of interest are made more invariant to changing imaging conditions and affine transformation by constructing a suitable descriptor for attaining more precise invariance properties. These descriptors define the region surrounding point of interest with some distinguishable property that helps in more accurate feature tracking results in a series of image under different affine transformations [Lin et al. 2009, Gauglitz et al. 2011].

Comparison between Marker based and Markerless Augmented Reality: Both the techniques for designing an AR system i.e. marker based AR and natural feature based (markerless) AR have their own advantages and disadvantages. Marker based approach is considered appropriate in cases where environment is very dense or cluttered i.e. the extracted information from the image scene involving color intensities or shape and structures of objects might be confusing due to various similarities, in such cases adopting an appropriate marker to track the position of virtual objects in the image scene seems beneficial. This approach is also adopted by many application developers due to the fact that the processing time taken to identify and track a well-recognized marker in an image scene is substantially low. AR systems based on markers, however are often not robust enough against environmental noise and these systems sometimes may even crash when markers are occluded. This approach may also require use of different markers in cases when the scene contains regions having patterns similar to the markers used, hence decreasing the receptiveness of AR system.

Markerless approach, on the other hand gives a flexibility to choose any part of the image scene as a target to superimpose virtual information resulting in improved applicability and robustness of AR systems for various applications.

Justification for selecting Markerless approach for the research: Markerless approach for designing an AR system incorporates properties that are suitable for distinctive range of applications and different scenarios.

- The second step in designing an AR system deals with integrating virtual and real world by using display technologies such as: 1) Optical-see-through: where the augmented output is viewed through a transparent screen at the actual scene. Optical-see-through techniques are used with head-mounted displays in form of beam-splitting holographic optical elements and in hand-held displays and spatial setups, augmented scene is reflected either using a planar screen or a curved screen. Optical-see-through displays tend to preserve the real-world resolution and are also cost effective and parallax-free i.e., it doesn't result in eye-offset due to camera positioning. 2) Video-see-through: is a more complex approach for viewing the augmented output because the system needs to evaluate camera position with respect to the user's location to offer synchronization between the real scene and virtual graphics. This AR display technology is one of the most adopted technology due to its favorable properties like, cost effectiveness, easy to use and it gives a wide adaptability of different real environments to achieve AR. Since real world is digitized, Video-see-through technology makes it possible to remove, add or alter real objects with much ease. This includes removing or replacing fiducial markers in the real scene and matching the brightness and contrast of virtual objects with the real environment parameters. The digitized real environment allows much accurate tracking of user's head movement for better registration. However, Video-see-through technology limits the field-of-view and user disorientation due to a parallax (eye-offset) which results from camera positioning at a distance from the viewer's true eye location, causing significant adjustment effort for the viewer.
- Interaction in Real time: AR system needs to process at a near frame rate so that it can superimpose information in real time and allow user interaction. When using AR for interactions, it is usually practical to have frame rate at least equal to six frames per second [Matas et al. 2004].

1.3 Augmented Reality System Design

Introduction: Augmented Reality system uses certain sophisticated technologies for estimating the accurate position and alignment of virtual objects that are to be blended in real environment. Figure 1.3 and Figure 1.4 represents the mechanism of two computer vision approaches, marker based and markerless approach, for attaining the desired augmented result of the virtual and real environment.

Explanation of Block Diagram: Figure 1.3 represents an example of marker based AR System building mechanism which uses a marker for estimating the pose and orientation of virtual object. In this approach, a predefined marker is tracked in subsequent image frames and in each frame the estimation of

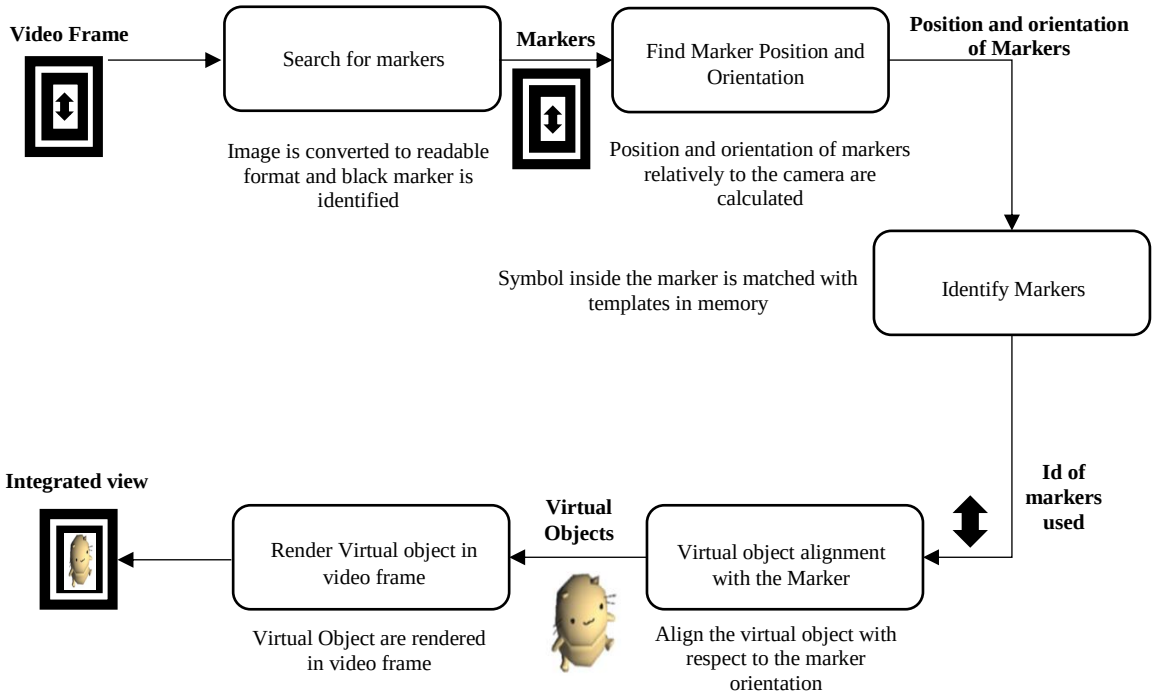


Fig. 1.3. Marker Based AR

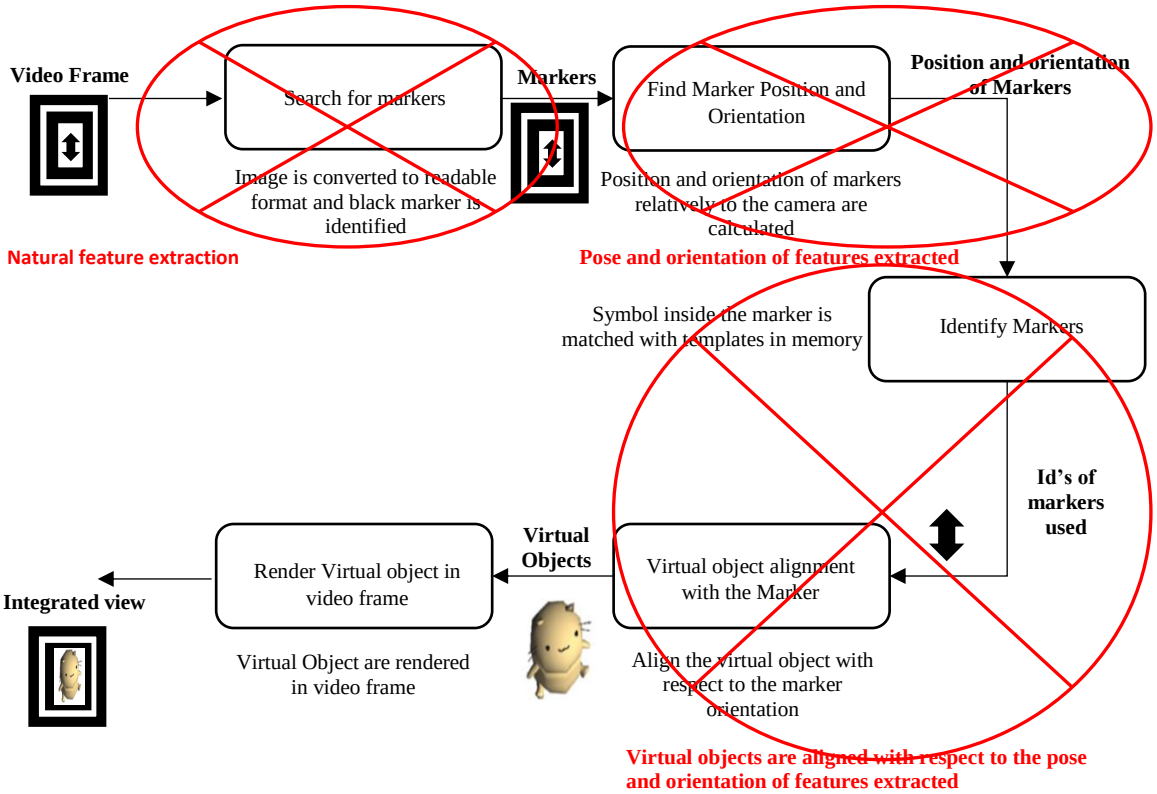


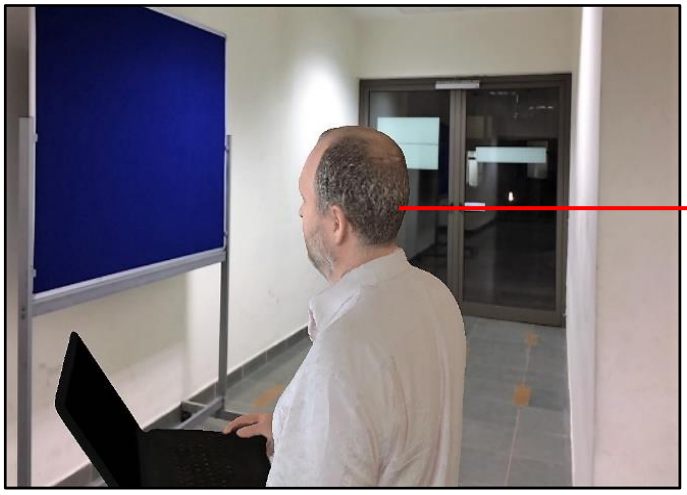
Fig. 1.4. Markerless AR

pose of virtual object is evaluated with respect to it. Figure 1.4 depicts the difference between a marker based and markerless AR system, where instead of searching the beforehand placed manmade markers in the image scene, natural features from the scene are extracted and accordingly, position and orientation of these extracted features are relatively calculated with respect to camera. As a result, alignment of virtual objects in the image scene is also evaluated with respect to the calculated pose and orientation of extracted features.

An example of Markerless Augmented Reality: Figure 1.5 represents an example of markerless AR system where a virtual man holding a laptop in his hands is rendered in the view of real scene without using any manually placed markers in the real scene beforehand.

Scene Description: Real scene contains an empty hallway with a blue background wheel notice board and a two-way glass door with a plain white square poster on left door and a plain white rectangle poster on right door.

(a) Original real Scene



Scene Description: Augmented scene augments a virtually displayed man with a laptop in his hand staring at the notice board to the real scene.

(b) Augmented View of the original scene

Fig. 1.5. An example of Markerless AR system

1.4 Image Registration Problem in Augmented Reality

Need of Image Registration: The process of building a markerless AR system involves two main modules: 1) Image Registration: It aims in proper alignment of virtual objects with real world coordinate system to give it a realistic view. 2) Rendering: integrates the virtual components with real image using the estimated pose and displays the augmented image on the display. Figure 1.6 represents a block diagram for a markerless AR system.

Requirements in terms of Accuracy & Time complexity: To be practical, feature detection approaches adapted for image registration are expected to operate in real time or in near-real time with high efficiency. Moreover, features detected in the image are ought to be robust to: scale change, viewpoint change and other affine distortions because pose estimation and further formulations done to build an AR system highly rely on these detected features [Yuan et al. 2006].

Desired Invariance to Imaging Conditions/Image Quality and Transformations: One more desirable property of a feature detection method being used for building an AR system is its invariance to imaging conditions and quality of images captured (or available) for processing. It increases the applicability of AR system designed for scenarios where a huge section of involvement comes from outdoor environment with different quality and types of images. Therefore, adopting a feature detection method that is invariant to imaging conditions like illumination change, blur change etc. and works well with even low quality images will surely be valuable.

Synchronization of video frame and the overlaid information: These feature detectors are also expected to process with near frame rate so as the synchronization between the video frame and the

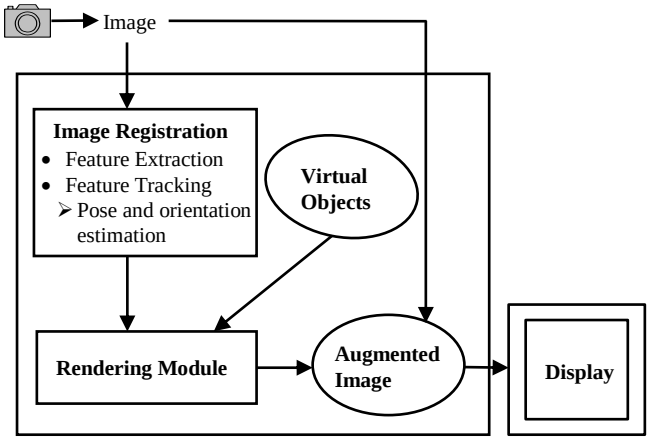


Fig. 1.6. Working Flow of Markerless AR System

overlaid information could be achieved seamlessly. Therefore to attain this objective, the feature detector used should be able to extract the most stable features in an image and should be able to track these features in subsequent image frames under unfavorable environmental conditions.

1.5 State-of-the-Art of Image Registration Methods in Augmented Reality Systems

Image Registration problem in AR aims in proper alignment of virtual objects with the view of real environment in 3D space. The task becomes more challenging when there are no manual markers used in the scene. For example, You and Neumann [You and Neumann 2001] focused on achieving stable and static 6DOF pose tracking by combining low frequency stability of visual sensors with high frequency tracking of gyroscope sensors. Li et al. [Li et al. 2005] aimed to improve the color-marker registration scheme to overcome the failure of capturing limited number of markers. A number of examples of computer vision methods for image registration are given by Chen et al. [Chen et al. 2010] and Hu et al. [Hu et al. 2005]. Lin et al. [Lin et al. 2006] also worked on the same ground with an objective of obtaining accurate and robust image registration results in an outdoor environment using phase correlation and Fourier-Mellin transform for a fixed-viewing-position AR system. Examples of hybrid registration combining the efficiency of both hardware and computer vision methods could be found in the research done by You and Neumann [You and Neumann 2001], Li et al. [Li et al. 2005] and Chen et al. [Chen et al. 2010].

Some examples of marker based computer vision registration methods detecting distinctive photometric features are; colored markers [Li et al. 2005], special shapes (eg. Squares [Li et al. 2007], dot codes [Hu et al. 2005], 2D barcodes [Azuma et al. 1999]) etc.

Since many years, researchers aim in extracting features or track patches of the natural image to achieve markerless registration [Lin et al. 2009, Jafari and Jarvis 2005]. These tracking approaches could be classified into two categories: First is the Feature based approach: i.e. tracking using local features e.g. Key-points, line segments and structure primitives [Li et al. 2005, Hu et al. 2005, Li et al. 2007]. Second is the Template based approach: which uses template correspondences (image patches) [Lin et al. 2009]. The advantage of feature based approach lies in its fast computation but it is difficult to model complex patterns by considering only the local features of the image.

Markerless Image Registration using natural features such as corner points, edges, planes etc. are also used for determining the camera pose estimation [Simon et al. 2000, Skrypnyk and Lowe 2004]. In order

to deal with general scenes, Simultaneous Localization and Mapping (SLAM) has been incorporated in robotics to track the camera pose while building a 3D map of the scene [Davison et al. 2007]. A more robust AR system proposed by Lee and Hollerer [Lee and Hollerer 2008] separates tracking and map building tasks instead of using SLAM. However, such a method needs manual calibration or a known-size object in order to provide initial metric scale information.

Optical flow algorithms are also used for efficient tracking [Lee and Hollerer 2008] and in order to achieve stable tracking results, landmark features are often used to overcome from complete tracking failures. Another way of locating landmark features is by using image patches referred to as templates [Yang et al. 2012]. Template matching for locating features or camera pose estimation can lead to ambiguous results in presence of repetitive patterns due to their fixed size. Therefore, multiple scale image patches could be used for robust results. In case of use of multiple scale image patches, invariant features that are consistent to scale and orientation are easily detected after working over multiple frames, but this process increases the computational complexity which makes the approach inappropriate for real time applications [Yang et al. 2012, Yang et al. 2013].

Over the years, AR system development has also been made using few different approaches. Giving a most recent example, an optical-see-through AR system developed by Wang et al. [Wang et al. 2017] aimed in overcoming three challenges of an AR system, i.e., registration, vision correction, and readability under strong ambient light using two liquid crystal lenses and a polarizer-free attenuator. The registration process in their approach is attained by using one of the liquid crystal lens to electrically adjust the position of the projected virtual image in the real environment.

1.6 Research Agenda

From the discussions in the earlier sections of this chapter, we observe that image registration plays a fundamental role in designing an AR system. The conventional image registration methods are generally expensive in terms of computational complexity to integrate with real time AR applications. Choosing an inappropriate method for feature detection also makes an AR system to suffer from incorrect alignment of virtual objects in the desired real environment [Wang et al. 2014]. Therefore, it is expected to develop a feature detection method for image registration which works well with extreme changing imaging conditions, hindered image quality, affine transformations and simultaneously it should be less expensive in terms of computational complexity. Also, due to lack of any prior research study that explained the performance of image registration methods for varying image quality and imaging conditions, there is a

need to compare the existing methods and use the data provided by the results to develop a new method or modify an existing method to work well under these imaging conditions.

Achieving the above mentioned objectives in designing an image registration method would solve the problem of incorrect calibration of virtual objects in real environment under varying conditions of 1) imaging conditions like illumination change, blur change, etc. 2) image quality, 3) affine transformations like scale change, viewpoint change etc. and would gain its receptiveness in real time AR applications and would also allow more accurate alignment of virtual objects in an AR system.

In the next chapter, these problems and research in these areas are investigated through literature review to lay down formally the objectives of the present research work.

1.7 Organization of the Thesis

A brief introduction to image registration, AR, need of image registration in AR and definition of various terms are given in this introductory chapter. The major problems in designing a markerless AR system, especially in image registration stage are summarized. Research trends and importance of crafting an appropriate image registration method for AR has been discussed in this chapter. The state-of-the-art image registration approaches are also discussed briefly.

Chapter 2 details the literature survey on various image registration methods proposed till date. It also includes AR history and evolution, discussing distinct image registration methods appropriate for building an AR system. This chapter also discusses image registration methods which can handle varying imaging conditions, affine transformations, varied image quality and noise in images.

Chapter 3 explains the relevant theory of various feature detectors, feature descriptors, varying imaging conditions and image quality metrics. Emphasis has also been given on the techniques used for comparing the performance of image registration methods. The interpretation and physical meaning of the various parameters is also explained. This chapter also explains the mathematical aspects of the theoretical concepts and gives a brief review of different transformations present in an image scene.

Chapter 4 discusses the effect of image quality and varying imaging conditions on image registration methods in two setups with different set of images and methods used for comparison. The comparative study detailed in this chapter involves the behavior analysis of image registration methods with respect to No-Reference and Full-Reference image quality assessment metrics.

Chapter 5 includes an improved model for No-Reference Image Quality Assessment model and a No-Reference Video Quality Assessment model based on frame analysis. The chapter proposes a Multi-Linear Regression (MLR) based model for No-Reference image quality assessment using three existing techniques and also presents few improved methods for estimating different distortions in an image frame.

Chapter 6 proposes an improvement in image registration using an improved implementation of maximally stable extremal regions for AR applications. This chapter discusses few existing drawbacks in image registration methods proposed till date and explains a procedure for handling some of the listed drawbacks using maximally stable extremal regions.

Chapter 7 summarizes few widely used feature description procedures and presents an improved feature descriptor for defining and sampling invariant regions surrounding the extracted keypoint. The proposed descriptor is based on local circular and elliptical sampling of image pixels and the chapter discusses detailed implementation of the proposed method.

Chapter 7 is followed by **Conclusions** of the work, **Specific Contributions** of the work done, **Critical Assessment of the Work & Suggestion for Future Research.**

Finally, **List of References** is appended which is followed by five **Appendices. Appendix A** contains the details of all the datasets used in this work. **Appendix B** contains implementation details of No-Reference Image Quality Assessment metrics, Full-Reference Image Quality Assessment metrics and Feature detectors. **Appendix C** contains complete listing of all the programs developed for developing a No-Reference Image Quality Assessment model and a No-Reference Video Quality Assessment model based on frame analysis. **Appendix D** contains implementation details for the proposed improvement in image registration using an improved implementation of maximally stable extremal regions for AR applications. **Appendix E** contains implementation details for the proposed feature descriptor for defining and sampling invariant regions surrounding the extracted keypoint.

At last, list of all the research papers, published, accepted or communicated for publication from this research work is given. This list is followed by a brief biography of the candidate (research scholar) and the supervisor.