CHAPTER

10

FEATURE DETECTION TECHNIQUES FOR MODEL-BASED TRACKING

Ajune Wanis Ismail, Nur Dinie Safirah Mohd Nazir, and Goh Eg Su

10.1 INTRODUCTION

Mixed reality (MR) elaborated on a virtual continuum from the real environment to virtual reality (VR). Augmented reality (AR) has classified the elements as either AR or augmented virtuality (AV) and looked at the tendency of how much both elements are between real and virtual objects. This is a taxonomy of MR that has been drawn by Milgram and Kishino (1994), as in Figure 10.1.

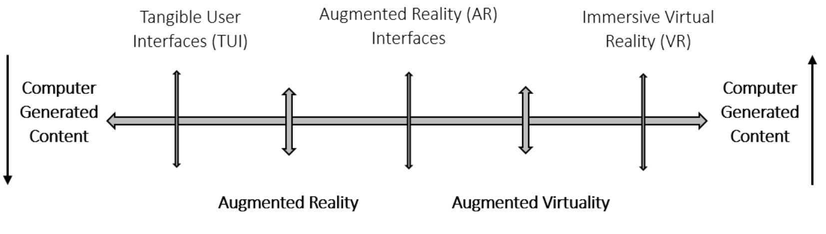

Figure 10.1 Taxonomy of mixed reality by Milgram and Kishino (1994)

In AR, the tracking, registration errors, and calibration are the common issues to be solved and improved by researchers. Registration error is an accurate alignment of the real and virtual objects. The illusion between the virtual and real objects exists in the real space and is severely compromised without accurate registration. The tracking of objects in the scene amounts to calculating the pose between the camera and the objects. Virtual objects can then be projected into the scene using the pose. Vision-based tracking determines the camera pose relative to real-world objects for closed-loop systems, which correct errors dynamically by using image processing.

Tracking is the most important element, and it is fundamental for enabling AR technologies. AR tracking techniques can be categorised into sensor-based, vision-based, and hybrid tracking techniques (Zhou et al., 2008). Vision-based tracking is one of the most widely used techniques that utilises image information to track the position and orientation of a camera. The ARToolkit was developed by Kato and Billinghurst (1999), who used computer vision (video or webcam) and markers to achieve AR. The marker is typically a rectangle with four corner points, which is sufficient for pose estimation. A 3×4 matrix is filled in with the video camera world coordinates relative to the card. After that, the matrix has been used to set the position of the virtual camera coordinates. Feature extraction in image processing refers to the process of transforming raw image data into a set of measurable and meaningful descriptors (called features) that can be used for tasks like classification, detection, recognition, or tracking.

10.2 UNDERSTANDING THE BASICS

Feature extraction in image processing follows a systematic pipeline, as presented in Figure 10.2, that begins with an input image, typically in colour. This image undergoes pre-processing to enhance its quality and consistency. The steps, like resizing dimensions, converting to greyscale, simplifying the data by removing colour, and filtering (e.g., Gaussian blur), help reduce noise. This stage ensures that the image is in optimal condition for further analysis. After pre-processing, feature detection identifies key points—distinct and repeatable patterns like corners, edges, or blobs—using algorithms such as Harris corner detection, Scale-Invariant Feature Transform (SIFT), or Oriented FAST and Rotated BRIEF (ORB). These points represent significant visual information in the image.

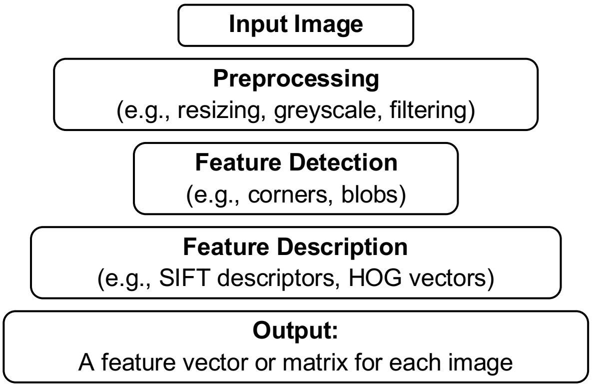

Figure 10.2 Basic systematic pipeline

Once key points are detected, feature description transforms each one into a numeric vector using descriptors like SIFT (128-dimensional), histogram of oriented gradients (HOG), or ORB. These descriptors encode the appearance around each key point, making them suitable for comparison across different images. The final output is a collection of feature vectors or matrices representing the image in a compact, discriminative form. These vectors are used in tasks such as image classification, object recognition, tracking, and AR, enabling systems to 'understand' and interpret visual data effectively.

Table 10.1 shows the type of features and their examples. In image processing, features can be categorised based on the type of information they capture. Colour features rely on colour space information, such as RGB or HSV, using techniques like colour histograms or average colour values to describe the image. While for the texture features, it focuses on patterns in surface or greyscale variations, with common methods including grey level co-occurrence matrix (GLCM), local binary patterns (LBP), and Haralick features. Shape features describe the outlines and geometric structure of objects, often represented using contours, HOG, or Fourier descriptors. Edge features capture boundary