CHAPTER II

Literature Review

2.1 Real-Time 3-D Pose Estimation System

In the early period, most of the work for real-time 3-D pose estimation used a binocular or stereovision. Tanaka, Maru and Miyazaki (1994) proposed a 3-D object tracking technique using active stereo vision system. The object and corresponding coordinate were extracted from the background for each camera (time delay or latency between cameras exist) and epipolar geometry was used to calculate the 3-D coordinate of the object.

To improve the visual information, multi-camera system will be used. Yonemoto, Arita, Matsumoto and Taniguchi (1999) developed a real-time coordinate capture system of a specified 3-D object based on multi-camera system using color marker. To improve performance of computation, the system implemented on a PC-cluster with network time protocol for PC-PC communication. Each camera connected to a dedicated PC. So that the real-time 3-D tracking was possible. However, this work did not concentrate on the tracking accuracy. The main purpose of the work is for tracking human motion behavior.

Garcia, Batlle and Salvi (2002) developed a trinocular stereovision system for real-time pose detection. Each camera embedded with a real-time image processing hardware to perform object labeling and noise filtering at video rate. Both 3-D position and velocity could be obtained with this method.

2.2 Camera Calibration

Camera calibration plays very important role for our application. Most of the vision-based systems for pose estimation of a scene require accurate prior known of system parameters, which can be estimated through a camera calibration process. The camera calibration process is based on the analysis of image feature of one or more views. A number of camera calibration methods have been proposed for the best result. They can be classified into two categories as the photogrammetric calibration and the self-calibration or auto-calibration.

2.2.1 Photogrammetric Calibration

The photogrammetric calibration is performed by observing a calibration pattern whose geometries in 3D space are known accurately. The 2-D coordinates, with correspondence 3-D data, obtained from each camera, are used to calculate the camera parameters. There are many calibration methods can be done very efficiently, such as:

Tsai' calibration method (1987) used monoview with coplanar or non-coplanar set of points, of the known pre-specify object, to compute camera parameters including the radial lens distortion using the projective geometry and Taylor's series expansion.

A three-step camera calibration method, proposed by Bacakoglu and Kamel (1997), used linear least-squares to approximate camera parameters in the first step. In the second step, Bacakoglu and Kamel develop the alternative formulation to obtain an optimal rotation matrix from approximated parameters. Then translational and perspective transformations were optimized based on the optimized rotation matrix. In the third step, non-linear optimization is performed to handle lens distortion.

Batista, Araujo and de Almeida (1999) proposed the iterative multi-step, explicit camera calibration method, which is based on iterative approach to avoid the singularities obtained by the calibration equations when monoplane calibration points are used.

Zhang (2000) proposed the calibration procedure based on known coplanar points in 3-D of the calibration object. The object was taken in different view points using the same camera. Using the homography, both intrinsic and extrinsic parameters can be found.

2.2.2 Self-Calibration

The self-calibration methods do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides constraints to intrinsic parameters. The correspondences between images, which are captured by the same camera in different view point, are sufficient to recover both intrinsic and extrinsic parameters. 3-D pose can be reconstructed up to a similarity. However, we cannot always obtain reliable results because there are many parameters to be estimated.

2.3 3-D Reconstruction

The 3D-reconstruction routine is needed in the 3-D pose estimation. In multiple cameras vision-based system for pose estimation, when number of image points (from two or more cameras used) of the calibration object are precisely known, as well as the intrinsic and extrinsic parameters of the calibrated cameras, the 3-D coordinate can be determined from the intersection in space of back projection rays. Each ray passes through the optical center and the known 2-D point in the image plane of the corresponding camera. These rays will intersect at the same point. Due to the presence of noise, these rays are not guarantee to intersect at a single point. There are some commonly-suggested methods to overcome this problem as:

2.3.1 Midpoint

This method compute 3D-point by minimizing the sum of the square distances of the 3D-point to each projected ray. However, this method strictly valid only in a Euclidean coordinate frame. Beardsley and Zisserman (1997) suggest an alternative method based on Quasi-Euclidean to find the average of midpoint of common perpendicular between any two rays. This method consumes less computation and acceptable result especially in Euclidean Frame, otherwise the error still exists but the result of reconstruction is better than original midpoint method.

2.3.2 Least-Squares and Iterative Least-Squares

This method uses less computation and gives high accuracy. For N cameras, the 2N linear equations are obtained from the relationship between camera model and points in the image planes. Lease-Squares method uses Singular Value Decomposition (SVD) or Pseudo-invert matrix to solve the 2N linear equations with 3 unknown to obtain 3D-point. This method has no geometrical meaning and its results vary with the weights upon its linear equations. Hartley and Sturm (1997) proposed an alternative method called iterative least-squares method. The original lease-square method is modified by adding weighting factor to the linear equations. The suitable weighting factor is adjusted each iteration. The result is more accurate but consumes more computation than the original lease-square method.

Liu et al. (2005) use least-squares method to reconstruct 3D-point from corrected image points. The first-order maximum likelihood estimation use to correct image points, which assumed a Gaussian noise distribution embedded in measurement. This method can be reconstructed 3D-point more efficient than both original least-squares method and iterative least-square method.

2.3.3 Bundle Adjustment

Bundle adjustment is the method to solve the problem of refining a visual reconstruction to produce jointly optimal 3D-structure and viewing parameter by using some optimization method such as Levenberg-Maquardt. There are many optimization methods used in bundle adjustment as shown in Triggs et al. (2000).

Hartley and Zisserman (2003) seek the maximum likelihood solution assuming that the measurement noise is Gaussian. They minimize the image distance between the detected image points and reprojected points.

Bartoli (2002) introduces an algorithm for bundle adjustment based on quasi-linear optimization to obtain 3D model from long image sequences.