Computers in Industry 70 (2015) 183–193

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Dynamic 3D surface reconstruction and motion modeling from a pan–tilt–zoom camera

Salam Dhou, Yuichi Motai *

Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form 10 February 2015
Accepted 17 February 2015
Available online 20 March 2015

Keywords:
Pan–tilt–zoom camera
Video stream reconstruction
Structure-from-motion
Target tracking

ABSTRACT

3D surface reconstruction and motion modeling has been integrated in several industrial applications. Using a pan–tilt–zoom (PTZ) camera, we present an efficient method called dynamic 3D reconstruction (D3DR) for recovering the 3D motion and structure of a freely moving target. The proposed method estimates the PTZ measurements to keep the target in the center of the field of view (FoV) of the camera with the same size. Feature extraction and tracking approach are used in the imaging framework to estimate the target's translation, position, and distance. A selection strategy is used to select keyframes that show significant changes in target movement and directly update the recovered 3D information. The proposed D3DR method is designed to work in a real-time environment, not requiring all frames captured to be used to update the recovered 3D motion and structure of the target. Using fewer frames minimizes the time and space complexity required. Experimental results conducted on real-time video streams using different targets to prove the efficiency of the proposed method. The proposed D3DR has been compared to existing offline and online 3D reconstruction methods, showing that it uses less execution time than the offline method and uses an average of 49.6% of the total number of frames captured.

1. Introduction

3D surface reconstruction has been integrated in camera-mounted computers for industry applications in recent years. Thousands of companies manufacture surveillance cameras. However, the expected software that can facilitate the analysis of the huge flow of video information is virtually absent. A useful idea would be to use 3D information from one off-the-shelf camera to extract the target motion and structure from a sequence of 2D images. Dynamic 3D reconstruction of structure and motion of targets is an important topic in industry [1–3]. It emerged in many recent industrial applications including real-time 3D image visualization [4], face detection [5,6], and 3D surface and curve reconstruction [1,7].

The advantage comes when the 3D reconstruction method uses less time and space to reconstruct the 3D objects from fewer frames, hopefully without incorporating a second camera [8]. 3D reconstruction using a pan–tilt–zoom (PTZ) camera is very promising due to its active imaging function, such as changing focus, pan–tilt head moving in network camera applications [9,10]. The work presented here specifically features a PTZ configuration so that the 3D object modeling of the target can be extracted from one single actively moving camera. In dynamic environments, a big need arises to combine the 3D reconstruction with tracking in which the 3D structure and motion are sequentially updated considering tracking information [11–13]. An active PTZ system can be useful for modeling because it uses the estimated PTZ motion to keep the target in the center of the image and updating the 3D model of the target. We propose a new method, D3DR, to recover the 3D motion and structure of the target, incrementally considering information from a single PTZ camera tracking the target.

The contribution of this paper is twofold: (1) proposing a dynamic 3D reconstruction method (D3DR) that dynamically estimates PTZ measurements of a camera tracking a target, and (2) proposing a keyframe selection strategy to select the frames that have a significant amount of motion to be used in the reconstruction process. Those new approaches simply update the 3D structure and motion of the target incrementally with fewer frames. We have specialized the imaging device as a latest PTZ

* Corresponding author. Tel.: +1 804 828 1281.
E-mail addresses: dhous@vcu.edu (S. Dhou), ymotai@vcu.edu (Y. Motai).

http://dx.doi.org/10.1016/j.compind.2015.02.005
0166-3615/

camera, and compared our proposed D3DR to those existing salient methods (called offline and online methods using the orthographic projection) in the identical pan–tilt–zoom setting.

Our overall proposed flow works as follows: the user selects an area inside a target by dragging the mouse to form a rectangular area. Then, the camera tracks the target area to keep it in the FoV of the camera with the same size. The keyframe selection strategy is applied to select the frames that show significant motion of the target to be used in the reconstruction process. The algorithm updates the target's motion and structure after every keyframe selected. Thus, the proposed method uses the tracking information immediately without the need to store growing matrices in memory.

The remainder of this paper is organized as follows: Section 2 presents a related study about 3D reconstruction. Section 3 reviews the existing offline and online reconstruction methods. Section 4 presents the proposed D3DR. The proposed method's performance is tested and the result analysis is shown in Section 5. Section 6 concludes the paper.

2. 3D reconstruction background

We present a brief reference background of 3D reconstruction methods that are relevant to our research. Section 2.1 reviews the problem of target tracking. Section 2.2 reviews 3D motion and structure recovery.

2.1. Target tracking

Feature points extraction and tracking is one of the most important parts of the reconstruction process. Optical flow can be used as a method for feature tracking. Yilmaz et al. [14] reviewed and classified the state-of-the-art tracking methods into different categories based on the object and motion representations used. Feature-based methods are developed to work in long image sequences for structure-from-motion problems [15,16] or for dynamic environments when both the camera and the target are moving [17]. A tracker should be able to detect and track the good feature points [18]. Kanade–Lucas–Tomasi (KLT) tracker [19] assumes the affine projection model and uses the least sum of squared difference (SSD) of pixel intensity between frames in the feature neighborhood to estimate new feature location. We choose the KLT tracker for our implementation because it has been demonstrated to be robust and reliable. Active PTZ network cameras are used for tracking a target as in [20–22].

2.2. 3D motion and structure recovery

3D reconstruction methods take advantage of information provided by a long sequence of images [23–25]. In [26], the method constructs 3D shapes of geometrical entities based on the iterative closest point (ICP) algorithm. Bozdagi et al. [27] proposed a method that estimates 3D motion and adapts a generic wire-frame to a particular speaker simultaneously within an optical flow based framework. Feature-based structure-from-motion methods including [28–30] depend on feature points selection and tracking in an image stream. The recovery and triangulation of 3D trajectories has been studied in [31,32], respectively. Online 3D reconstruction has advantages over the offline 3D reconstruction methods when used in real systems [11–13]. Automatic reconstruction methods exist to recover motion and structure from monocular image sequences [33–35] or using feature-based structure-from-motion approaches [28,36]. Extended Kalman filter is used to recover motion and structure from an uncalibrated video [37]. Distributed and scalable volumetric architecture for reconstructing arbitrary structures in real time are used as in [38]. This architecture consists of acquisition nodes to reconstruct partial models from multiple views and a master node to merge those partial models.

3. Existing 3D reconstruction studies

We present existing 3D reconstruction methods directly relevant to our research. Section 3.1 shows the offline reconstruction using the factorization method. Section 3.2 shows the online reconstruction from a moving camera using the sequential factorization method.

3.1. Offline 3D reconstruction from a stationary camera using the factorization method

For the recovery of 3D structure and motion, we present Tomasi and Kanade's factorization method [29]. The input of this method is the feature points correspondences provided by the KLT tracker. The KLT tracker used in this paper rejects the outliers and chose the "good" features to track. We assume a sequence of F images taken for a moving target in front of a PTZ camera. The orthographic projection model assumes that projection rays are parallel to the camera's optical axis.

Fig. 1 shows the systems of reference for the image, target, and camera. For frame f, the camera orientation is described by the unit vectors i_f, j_f, and k_f. The distance between the camera origin and the fixed world coordinate is represented by the translation vector T. Each point $s_p = [X, Y, Z]^T$ in the fixed world coordinate is located by KLT at the image position $p_{f,p} = (u_{f,p}, v_{f,p})$ in image frame f and located at feature point position $p_{f-1,p} = (u_{f-1,p}, v_{f-1,p})$ in frame $f-1$.

Given a video stream, suppose that we have tracked P feature points over F frames. We then have the image coordinates: $p_{f,p} = \{(u_{f,p}, v_{f,p}) : f = 1, \ldots, F,\ p = 1, \ldots, P\}$. The corresponding feature points $p_{f,p} = (u_{f,p}, v_{f,p})$ are written in a measurement matrix $W : 2F \times P$:

$$W = \begin{bmatrix} u_{11} & \cdots & u_{1P} \\ \vdots & \ddots & \vdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ \vdots & \ddots & \vdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix}. \tag{1}$$

The registered measurement matrix W^* is computed by subtracting from W the mean $x_f = \sum_{p=1}^{P} u_{fp}/P$ and $y_f = \sum_{p=1}^{P} v_{fp}/P$ of all $u_{f,p}$ and $v_{f,p}$ at a frame f respectively:

$$W^* = \begin{bmatrix} U \\ V \end{bmatrix} - \begin{bmatrix} x_f \\ y_f \end{bmatrix} = MS \tag{2}$$

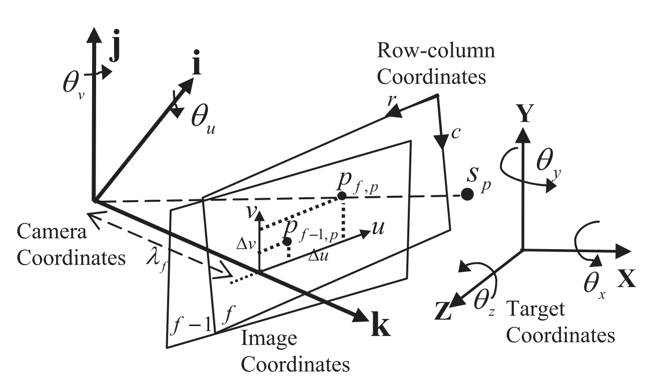

Fig. 1. The coordinate system showing the geometric relation between frame $f-1$ and frame f.