Effectual Method FOR 3D Rebuilding From Diverse Images

Carlos Wai Yin Leung, B.E. Hons
University of Queensland, Australia

Abstract

This thesis explores the problem of reconstructing a three-dimensional (3D) scene given a set of images or image sequences of the scene. It describes efficient methods for the 3D reconstruction of static and dynamic scenes from stereo images, stereo image sequences, and images captured from multiple viewpoints. Novel methods for image-based and volumetric modelling approaches to 3D reconstruction are presented, with an emphasis on the development of efficient algorithm which produce high quality and accurate reconstructions.

For image-based 3D reconstruction a novel energy minimisation scheme, Iterated Dynamic Programming, is presented for the efficient computation of strong local minima of discontinuity preserving energyy functions. Coupled with a novel morphological decomposition method and subregioning schemes for the efficient computation of a narrowband matching cost volume, the minimisation framework is applied to solve problems in stereo matching, stereo-temporal reconstruction,motion estimation, 2D image registration and 3D image registration. This thesis establishes Iterated Dynamic Programming as an efficient and effective energy minimisation scheme suitable for computer vision problems which involve finding correspondences across images. For 3D reconstruction from multiple view images with arbitrary camera placement, a novel volumetric modelling technique, Embedded Voxel Colouring, is presented that efficiently embeds all reconstructions of a 3D scene into a single output in a single scan of the volumetric space under exact visibility. An adaptive thresholding framework is also introduced for the computation of the optimal set of thresholds to obtain high quality 3D reconstructions. This thesis establishes the Embedded Voxel Colouring framework as a fast, efficient and effective method for 3D reconstruction from multiple view images.

1. Introduction

1.1 Problem Statement

The human vision system obtains information about the world through planar images that are formed on the retina of each eye. Depth and three dimensional structures are perceived through the analysis of the stereo images captured from the left and right eyes. What appears to be so natural and simple for human beings however has proven to be surprisingly difficult for computers. The aim of computer vision is to mimic human's ability in analysing visual information and to enable computers to *see* like we do. This thesis focuses on one of the many challenges in computer vision — the process of reconstructing a three-dimensional (3D) scene and the recovery of the depth information given a set of photographs. In particular, this thesis explores the 3D reconstruction of static and dynamic scenes from stereo images, stereo image sequences, and images captured from multiple viewpoints. Furthermore the design of efficient algorithms will be a major consideration as well as the accurate reconstruction of the 3D scene.

1.2 Motivation

1.2.1 Application Areas

The 3D reconstruction of a scene from images alone has many useful applications. In the entertainment industry, it has been widely applied to aid in the process of moving making. A scene with a large number of horses galloping in a field for example, can be achieved by firstly reconstructing the dynamics of one horse galloping in the field. New views of the horse can then be synthesized from the 3D horse model obtained, and the scene can be created by inserting as many horse models into the scene as desired. Without the aid of 3D reconstruction, computer graphics artists would need to spend many tedious hours of CAD-modelling while often faced with the problem of a lack of photo-realism when the objects are rendered. Other techniques such as attaching markers onto actors and recreating their 3D models by tracking these markers have also been explored. These markers however are often inconvenient and uncomfortable to actors. Not only do they disrupt the performances of the actors, photo-realistic 3D models based upon markers are also extremely difficult. The ability to create 3D reconstructions using images alone is therefore a very attractive alternative. 3D reconstruction from multiples views in particular has found many applications in creating movies, computer games and animations.

3D reconstruction from images is also widely applied in the medical industry. It has been used to create models of a whole range of organs, as well as brains and even teeth. Other application areas include body motion modelling, teleconferencing, robot navigation, object recognition, surveillance, and surveying such as the modelling of terrain and buildings. The number and quality of the images supplied generally depends on the application. Robots and terrain modelling for example tend to capture images using stereo cameras, while human body modelling can usually allow for a multiple camera setup. In recent years there has also been interests in implementing 3D reconstruction algorithms in hardware, so that embedded systems can begin to include 3D modelling and depth analysis capabilities into their applications. Despite the many applications however, there exist several opportunities for research in order to improve the 3D reconstruction process and results obtained. This thesis presents several advancements in the field of 3D reconstruction, introducing algorithms which are efficient while computing high quality and accurate results.

1.2.2 3D Reconstruction from Images

One of the key challenges of computer vision is the reconstruction of the 3D scene given only a set of planar images. A wide variety of algorithms and methods have been studied in the past. These

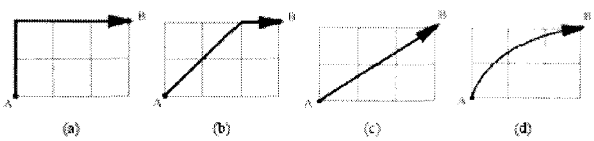

Figure 1. An example of a shortest path between points A and B using a (a) 4-connected discrete space, (b) 8-connected discrete space, and (c) the continuous solution. (d) A different shortest path may result if the grid is weighted.

Local algorithms refer to those methods which consider only local optimisation. These include for example methods which consider only local neighbourhoods or smoothness within a local region. Optimal methods on the other hand consider global optimisation and focus on the computation of optimal solutions. Dynamic programming for example computes the global shortest path across a trellis which will be described in Chapter 2. Problems which can be formulated into a graph cut framework for example, can be solved optimally by computing the maximum flow which yields the minimum cut. However there are a set of problems which are NP-hard that cannot always guarantee optimality, dependent upon the size of the problem and type of complexity. In cases where optimality cannot be guaranteed, one popular approach is to break the NP-hard problem into a set of subproblems which can be computed optimally. The larger the set of solutions considered by each subproblem, the closer in general is the approximation to the optimal solution.

Discrete methods solve problems based on a discrete modelling of the space; while continuous methods use a continuous model. Images are discrete because they are tesselated into a finite set of values. Continuous solutions represented by images therefore become discrete descriptions of the continuous space. But note that the solution is still continuous. The difference between discrete and continuous methods is illustrated in Figure 1. Consider the computation of the shortest path between A and B assuming uniform vertex and edge costs. For a 4-connected discrete space, the solution is shown in Figure 1(a). Using an 8-connectivity, the solution improves to Figure 1(b). Discrete methods are always limited by the edges, and are therefore always grid biased. The shortest path is illustrated in Figure 1(c) for the continuous case.

Further consider an example where the hortest path is the solution in Figure 1(d) with weighted grids. How close the result generated by a continuous method is to the true solution depends on the number of grid points. Since continuous solutions are represented by discrete descriptions of the continuous space, as the number of grid points increase, the continuous solution approaches the true solution. For discrete methods however, increasing the number of grid points does not draw the solution closer to the true solution because it is limited by the connectivity of the grids.

2. Preliminaries

This chapter presents background theories and concepts related to this thesis. A brief introduction to multiple view geometry is presented in Figure 2 Camera models and calibrations are discussed, along with the fundamentals of epipolar and projective geometry. One of the major focuses of this thesis is efficiency. Figure 2 describes efficient tools for computations involving shifting windows. The box filtering technique is an efficient method for computing the sum of the values within a shifting window. The application of grey-scale morphology is an efficient tool for computing the maximum or minimum value within a shifting window. Figure 2 describes the dynamic programming algorithm which forms for the basis for the Iterated Dynamic Programming algorithm, which is presented in 2.1.

2.1 Multiple View Geometry

The 3D reconstruction problem is a fundamental and heavily studied area in the field of computer vision. Previous research has focussed on problems such as the recovery of the epipolar geometry between two stereo images, and the calibration of multiple camera views. These research studies lay the foundation for current work in dense stereo reconstruction and volumetric modelling. In this section a brief introduction to multiple view geometry is presented. The camera models and projection theory described in the following sections are used throughout this thesis.

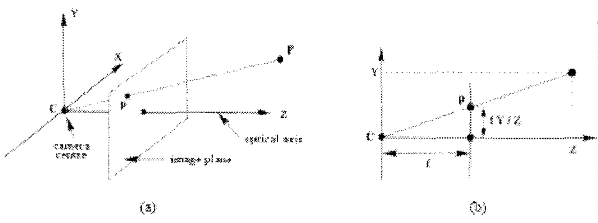

Figure 2. (a) The pinhole camera model. (b) A side view of the pinhole camera model.

2.2 Methods forWindowing Operations

This section presents tools that are useful for performing analysis over selected regions of an image. Section 2.2.1 presents a brief overview of window-based matching techniques, which is required in dense stereo reconstruction. It is also very common to perform summations over a windowed region. It is observed that when computing the sum for a shifting window, there are overlapping regions which result in redundant computations which can be minimised given an efficient algorithm. Figure 2 presents the box filtering technique for efficient summation of values within a shifting window. It is also common to compute the maximum or minimum value within a shifting window. Therefore in Figure 2 an efficient method to obtain the maximum or minimum value within a shifting window using grey-scale morphology is presented.

3. Efficient Methods for Matching Cost Computation

Image-based 3D reconstruction and problems involving finding correspondences can be conceptualized and divided into two major stages. The first stage involves the computation of a set of matching costs given a search range, while the second stage considers the selection of the optimal matches given a set of criteria. It is therefore important in the construction of an efficient 3D reconstruction system to consider methods to efficiently compute the matching cost volume. This chapter focuses on the development of efficient algorithms to minimise the computational time of the first stage. Accuracy is not considered because in the computation of the matching cost, there can only be one exact solution corresponding to the chosen matching metric. Different solutions and the optimisation for the best reconstruction occur in the second stage which is considered in this paper.

In this paper describes the applic ation of the box filtering technique to efficiently compute matching costs in a rectangular region. The equations for the matching metrics which are used throughout the thesis are presented, along with specialised implementations of these equations for minimised computational load. In this paper describes the coarse to fine approach and upsampling scheme. The coarse to fine approach is applied in order to limit the searchable disparity range to a restricted region of values at each resolution. Although the amount of matching values to compute is reduced, efficient computation of the matching costs in a narrowband is non-trivial. This chapter presents two novel algorithms to solve this problem.

In this paper a novel algorithm, quadtree subregioning (QSR), is presented to efficiently analyse the narrowband. A review of previous methods is presented, detailing areas where improvements can be made in order to further increase the efficiency of matching costs computation within narrowbands. The QSR algorithm is described and it will be demonstrated how QSR computes the optimal quadtree partitioning. The proposed method will be applied to the stereo matching problem, presenting results and timings to illustrate the efficiency of the QSR subregioning scheme.

A second algorithm for the computation of matching costs in a narrowband is then presented in this paper. The runlength filtering (RLF) scheme analyses the narrowband from a different perspective to previous methods. Instead of partitioning the restricted search range into subregions and computing the matching costs within these partitions, RLF encodes lines of the narrowband into runlengths and computes matching costs for the tightest bound of the narrowband. A thorough description and the implementation details of the RLF algorithm will be presented, along with timings to demonstrate the efficiency of its application.

4. Efficient Methods for Energy Minimisation

Given an energy function, a variety of algorithms have been proposed in literature which solve for the minimum energy. By formulating computer vision problems into an energy minimization framework, energy minimisation techniques can be applied to compute likely solutions. The 3D reconstruction problem can similarly be formulated into an energy minimisation problem by designing an energy function whose minima corresponds to good reconstructions. Such an energy function will need to incorporate the matching likelihood between image primitives as well as additional constraints such as regularisation. This implies that other problems which involve solving for the optimal matching between images, such as stereo-temporal reconstruction, motion estimation and image registration can similarly apply energy minimisation techniques to compute its solutions.

In recent years high quality solutions to the stereo reconstruction problem have been achieved through the application of energy minimisation techniques. By incorporating global constraints and structural information into the stereo matching process highly accurate disparity maps have been computed. Many of these techniques can be classified into methods based on dynamic programming and methods based on graph cuts. Dynamic programming based techniques can locate the path of minimum matching cost for each scanline of the image. While this allows efficient optimal solutions for each scanline, interscanline inconsistencies pose problems. Graph cut methods on the other hand consider the minimisation of a two-dimensional energy function and do not have problems with dimensional bias. Graph cut methods however are known to be computationally expensive requiring orders of magnitude greater computation than dynamic programming techniques.

This chapter presents a novel energy minimisation scheme, Iterated Dynamic Programming (IDP). IDP is a fast energy minimisation scheme based on dynamic programming however immune to the interscanline problem. Retaining the benefits of a dynamic programming framework, IDP produces high quality results for significantly reduced computation compared to graph cut approaches. This thesis proposes IDP as a new alternative for minimising general energy functions of the form to be described in this paper. While IDP can be applied to solve a variety of image matching problem which will be described in this paper, it is applied to the stereo reconstruction problem in this chapter in order to demonstrate the effectiveness and efficiency of IDP.

In this paper the formulation of the stereo reconstruction problem into an energy minimization framework will be presented. In this paper previous research in energy minimisation is described. In this paper will then describe in detail the IDP algorithm and demonstrate its application to the stereo reconstruction problem. A novel application of morphological decomposition for the fast computation of dynamic programming will also be presented. This is achieved by minimising the number of minimum operations in the forward sweep of the dynamic programming algorithm. Its application further reduces the computational time of the IDP algorithm. To support the theoretical explanations, results will be presented In this paper to demonstrate the efficiency and accuracy of the IDP algorithm. The high quality stereo reconstructions obtained using IDP will be presented along with timings to show the reduced computational time required by the algorithm. An extensive comparison to existing methods is provided in order to establish IDP as a competitive energy minimisation scheme.

5. Applications of Iterated Dynamic Programming

This chapter presents the application of the Iterated Dynamic Programming (IDP) energy minimization scheme presented in this paper to solve a variety of computer vision problems which involve finding correspondences. In this paper formulated the stereo reconstruction problem into an energy minimisation framework and applied IDP to compute a strong local minimum of a discontinuity-preserving energy function. Results was demonstrated to show the ability of IDP to efficiently compute high quality stereo reconstructions. In this chapter different matching problems are similarly formulated into an energy minimisation framework and IDP will be applied to solve these problems. The aim is to establish IDP as an efficient energy minimisation scheme suitable for a variety of computer vision applications.

In this paper will showcase a wide selection of stereo reconstructions computed using IDP. In this paper IDP is applied to the stereo-temporal reconstruction problem Given a stereo image sequence, the IDP energy minimisation scheme is adapted to optimise a three-dimensional domain space with a one-dimensional disparity search range. The application of IDP to solve motion estimation problems will then be presented in In this paper. While optical flow methods have been widely studied to solve the motion estimation problem, an energy minimisation approach using IDP is presented to efficiently compute solutions. In motion estimation IDP optimisation considers a two-dimensional domain space with a two-dimensional disparity search range. A computer vision problem that is similar to motion estimation is 2D image registration, which similarly considers a two-dimensional search range within a two-dimensional domain. In this paper IDP will be applied to the 2D and 3D image registration problem. The challenge of 3D image registration is the size of the problem since it involves a three-dimensional search range in a three-dimensional space.

Consider the matching problem as a mapping from one dimensional space to another described by the dimensions of the domain and disparity search ranges. The mappings considered by each of these applications are:

A common characteristic of each of these problems is that the aim is to assign each pixel with a matching pixel from another image. For each pixel, one searches within a defined neighbourhood and decides on the best choice of disparity based on a set of constraints. Therefore each of these problems can be conceptualised as a labelling problem. A matching cost volume can be constructed which encapsulates all likelihoods of matching and an energy function defined which describes how well a certain configuration matches the set of constraints. Since each of these problems can be formulated into an energy minimisation framework, IDP can be applied to obtain solutions.

In this paper a large moves analysis of IDP was presented. The size of the large moves considered for each of the applications described in this chapter was also presented. The following sections will describe in detail the formulation of IDP to perform large moves optimisation for each of these applications. Results will be presented to demonstrate the effectiveness and efficiency of IDP as an energy minimisation scheme suitable for a variety of computer vision problems. Since a matching cost volume will be computed for each these applications, all of the efficient cost computation methods described in Chapter 3 will similarly be applied.

6. Efficient Methods for Volumetric Modelling

The 3D reconstruction of a complex scene from multiple images is a fundamental problem in the field of computer vision. Given a set of images of a scene captured from arbitrarily placed cameras, the goal is to recover the unknown 3D structure using these images and knowledge of the camera geometry. In previous chapters image based 3D reconstruction methods was explored. These methods infer depth information through matching corresponding primitives between images. By the process of triangulation the 3D location of matched objects can be recovered, as depicted in Figure 3 While image based methods have produced excellent results for two view analysis such as stereo reconstructions, the extension to consider multiple views is non-trivial. Especially since image based methods depend on a reference image, the corresponding matches are generally associated to a disparity function of the pixel d(x, y). This implies that the 3D reconstruction is not completely 3D, but a 2.5D reconstruction of the scene. The modelling of occlusions is complicated and there is not a standardised and widely accepted framework for modelling occlusions for image based methods. Occlusions is currently one of the major focus of research. In recent years volumetric modelling methods have emerged as a popular alternative to image based methods for 3D reconstruction. Its main advantages lay in the ability of the volumetric framework to explicitly model occlusions and consider multiple views. A major disadvantage however which prevents the wide application of volumetric techniques is the extremely high computational load required.

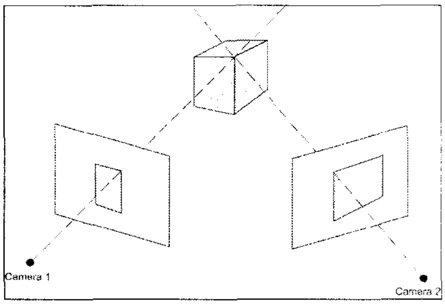

Figure 3. The 3D location of matched imaged points can be recovered through triangulation.

This chapter presents several advancements in volumetric modelling methods. The chapter begins with a presentation on previous work in volumetric modelling in this paper. Background information on how volumetric methods analyse multiple views and model occlusions will be discussed. In this paper will then identify current problems in volumetric methods and areas in which improvements can be made. It will define the problem and outline the issues this thesis is trying to solve in the areas of efficiency and accuracy. In this paper presents one of the three key properties, a water-tight visibility model, which have been identified as necessary for volumetric modelling algorithms. In this paper will describe two additional properties, a monotonic carving order and causality, which enable the derivation of the Embedded Voxel Colouring algorithm (EVC). While previous volumetric techniques require several runs of the computationally expensive algorithm to decide on the optimal choice of threshold, EVC removes the need for *a priori* threshold selection. EVC embeds the reconstructions of all possible thresholds into one output volume in a single run of the algorithm.

Given the embedded volume, one can select a threshold *posteriori* to obtain the desired choice of 3D reconstruction. In many cases however, a single threshold does not satisfactorily recover the 3D scene and adaptive thresholding is required. In this paper will describe an adaptive threshold selection algorithm which computes the optimal choice of thresholds using globally minimal surfaces. In this paper then describes in detail the implementation details of EVC and the adaptive thresholding framework. Results demonstrating the quality and speed of the EVC algorithm and adaptive thresholding process will be presented in this paper.

7. Conclusions and Future Work

This thesis presented novel energy minimisation based and volumetric modelling methods for the efficient computation of high quality 3D reconstructions from multiple view images and image sequences. The focus of this thesis have been on the development of novel efficient algorithms which produces high quality and accurate reconstructions of 3D scenes. As computing power increases, the computational times required for efficient algorithms will similarly improve. In this thesis, algorithms have been presented which reconstruct 3D scenes from stereo images, stereo video sequences and multiple view images. The Iterated Dynamic Programming (IDP) scheme was presented which can efficiently minimize an energy function. Coupled with a novel morphological decomposition method to efficiently compute the minimal path during each optimisation step, and a quadtree subregioning (QSR) and runlength filtering (RLF) scheme to compute the matching cost volume, the IDP framework was applied to solve problems in stereo matching, stereo-temporal reconstruction, motion estimation, 2D image registration and 3D image registration. In the case of 3D reconstruction from multiple view images, a novel volumetric modelling technique have been presented that removed the need for *a priori* threshold selection. By identifying three key properties essential to volumetric modeling algorithms, the Embedded Voxel Colouring (EVC) scheme was developed which efficiently embeds all reconstructions of the 3D scene into a single output in a single scan of the volumetric space. An adaptive thresholding framework was also presented which computes the optimal 3D surface with respect to the embedded output, producing high quality 3D reconstructions of the 3Dscene.

This thesis has demonstrated the efficiency and high quality reconstructions computed by the novel 3D reconstruction methods developed herein. Detailed analysis and comparisons with existing algorithms have shown that the proposed techniques can produce strongly competitive results at much reduced computational time. The algorithms presented in this thesis can be applied to efficiently solve 3D reconstruction problems and computer vision problems which involve finding correspondences across images.