%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\chapter{Discussion}

\label{ch:discussion}

This dissertation presents three works for the problems in static scene reconstruction and dynamic object reconstruction. In Chapter \ref{ch:patchmatch}, we propose a framework of joint view selection and depthmap estimation. The experiments on large Internet collected photos demonstrates its efficiency and robustness. In Chapter \ref{ch:jost} and Chapter \ref{ch:video\_l1}, we solve the problems of dynamic object reconstruction from unstructured images and unsyncthronized videos, respectively. In solving these two problems, our main effort is on 3D reconstruction without sequencing information. We show effectiveness of the approaches by testing on synthetic and real datasets. Next, we discuss the possible extension of our works, and the potential future research directions.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\section{Future work}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\subsection{Extensions to PatchMatch-based joint view selection and depthmap estimation}

\label{sec:patchmatch\_extensions}

Though our method significantly outperforms existing methods on Internet collected photos \cite{Goesele07}, and achieves the state-of-the-art accuracy on standard datasets collected under controlled lab environment \cite{Strecha08}. The accuracy of the method can be further improved by incorporating some standard techniques into our framework. Next, we discuss each of the techniques for possible accuracy improvement.

In the method, we use the frontal-parallel planes to warp color patches into other images for color consistency check. It has been shown the plane orientation affects the reconstruction accuracy \cite{Gallup07,FURUKAWA\_PAMI2010}. Ideally, the plane orientation should be close to the real surface normal, which is unknown before reconstruction. To address this issue, the surface normal can be included as unknown variables in our framework. Specifically, the normal directions are propagated to the neighboring pixels in addition to the depth \cite{patchMatchStereo1}. This scheme has a potential to further improve reconstruction quality on the regions having large angles with the camera viewing directions (e.g. the ground), but increases the computation complexity.

Another issue relating to the color patch arises if the pixels in a patch cover scenes of significantly different depths, which typically occurs at the boundary of object surfaces.

When compare color consistency between two pixels, current local methods (\ie no smoothness term in the depthmap) use the whole patches around the pixels to improve the robustness.

The method present in Chapter \ref{ch:patchmatch} uses NCC as a metric to measure the color consistency, where each pixel in the patch contributes equally for the measure.

However, this is likely to produce swollen/fat boundaries in the depthmap, since the usage of a plane for patch warping assumes all pixels in the patch have the same depth and normal, and this assumption breaks at the boundary of object surfaces. Therefore, when measuring the color consistency between two patches, the pixels with the same depth as the central pixel should be given more weights than other pixels. To achieve this, one effective solution is using adaptive weights for each pixel in the patch based on color similarity and geometric proximity relative to the central pixel \cite{Yoon06adaptivesupport\_weight}.

Another extension to our work is to handle cameras with small baselines. In stereo methods, small baselines usually lead to unstable and inaccurate results. Since the large set of Internet collected photos is typically taken at certain spots of interest, it is very likely some of the images have very small or zero baselines. Our framework select images based on color consistency, and the images with small baselines will be selected since the color consistency are always good regardless of estimated depth. To address this issue, the angles of two viewing rays given a depth hypothesis should be computed to prevent invalid triangulation \cite{Gallup08}. Specifically,when estimating the depth of a pixel, if the angle of two viewing rays is below a threshold, the corresponding source image is eliminated.

Though great advance in depth estimation has been achieved, there are still some important problems remain unsolved. One case in point is depth estimation for homegenerous color regions.

Most current methods estimate depth by checking color consistency, and none of them can successfully address this issue.

To handle this problem, I believe it is necessary to incorporate the semantic knowledge of the scene rather than to just rely on low-level colors. This inevitably requires introducing the machine learning techniques into stereo problems. However, the incorporation of camera parameters into machine learning frameworks has another issue, since the testing data and training data may have different camera parameters. Though there are many single image depth estimation based on supervised machine learning \cite{Hoiem\_CGRAPH2005,Saxena\_IJCV2008,eigen2014depth,Liu2014,zhuo2015indoor}, incorporating this technique into multiview stereo methods for accurate depth estimation still has long way to go.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\subsection{Extensions to JOST}

The method presented in Chapter \ref{ch:jost} uses object detection output as features, and the object lies along the viewing ray passing the features. However, our method cannot handle outlier detection (false alarm), and the outliers may stop the method from finding the correct trajectory. One way to manage this problem, as is done in the paper, is to raise the detection threshold to suppress the false alarm rate (at the cost of increasing miss detection). Another possible way is to embed our method in a RANSAC framework \cite{Hartley2004} to remove outliers. Specifically, the subset of sampled detections is used to triangulate the trajectory, and count the number of remaining detections censuses with the trajectory (\ie~inliers). Repeat this process to find the trajectory with the largest number of inliers. However, this scheme is computational intensive if the ratio of outliers is large, considering that running trajectory triangulation given a subset of detections is time-consuming.

Efficiency is another issue for our approach. In our method, the nonconvex problem is solved in a discrete-continuous scheme, and the discrete step involves solving a NP-hard GMST problem. The efficiency of solving a GMST problem can be attained by reducing the complexity of the K-partite graph. A complete K-partite graph has all the nodes in one set connecting to all the nodes in another set. The computation complexity of GMST given the graph will be lowered down if the number of edges and nodes is reduced. To simplify the K-partite graph, if given a prior knowledge that two specific detected objects are farther away in 3D space, then all the edges connecting the associated two sets of nodes can be safely removed. Moreover, the size of each object detection window can be used to roughly estimate the tight depth range of the dynamic object, which helps reduce the size of each node set and hence the number of edges.

\subsection{Extensions to dynamic object reconstruction from unsynchronized videos}

The method presented in Chapter \ref{ch:video\_l1} requires a static background scene present in the image so that structure from motion can use it for camera registration. However, the crowd sourced data may have the dynamic object as main focus and lack the content of the background scene. This comes an open question of how to register cameras given non-current captures of dynamic objects. Considering the importance and difficulity of this problem, it can be a possible future research direction.

Moreover, to the best of our knowledge, the work is the first self-representation framework for dynamic object reconstruction. That is, each temporal shape can be represented by a linear combination of a few other shapes given the smooth motion of dynamic objects. This self-representation constraint has potential to be applied in the NRSFM problems. Comparing to most of the existing works, where the assumption that any shape is a linear combination of $K$ shape bases is applied, our self-representation constraint is more intuitive and possibly lead to better reconstruction results.

% obtaining correspondences.

% possible dense reconstruction.