Abstract of thesis entitled:

Robust Multi-view Video Synopsis and Panoramic Navigation

Submitted by ZHANG, Zhensong

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in August 2018

With the rapid growth of the multi-view videos, effective and efficient organization and browsing of such contents bring about research challenges as well as opportunities. It is becoming an increasingly important issue to provide users with immersive and efficient multi-view video browsing experience. Panoramic display and video synopsis are promising methods for solving this issue. However, existing large-view panoramic display methods are usually based on computational intensive 3D reconstruction, and multiple synopsis outputs are difficult for users to comprehend. To tackle these problems, in this thesis, we focus our research on robust and efficient approaches in street view interpolation, dynamic video stitching for panoramic display, and compact multi-view video synopsis, which can be widely applied for multi-view surveillance videos, large-scale displays, and mobile cameras applications.

Wide-baseline street view interpolation is useful but very challenging. Existing approaches either rely on heavyweight 3D reconstruction or computationally intensive deep networks. We present a lightweight and efficient approach that utilizes homography computing and refining operators to estimate piecewise smooth homographies between
input views. We combine homography fitting and homography propagation based on reliable and unreliable superpixel discrimination. Our framework dramatically increases the accuracy and robustness of the estimated homographies. We integrate the concepts of homography and mesh warping and propose a novel homography-constrained warping formulation which enforces smoothness between neighboring homographies by utilizing the first-order continuity of the warped mesh. This further eliminates small artifacts of overlapping, stretching, etc. The proposed method improves the state of the art and demonstrates that homography computation suffices for multi-view interpolation. Our experiments on city and rural datasets validate the efficiency and effectiveness of our approach.

Video stitching produces a panoramic video with a large field of viewing, which can essentially enhance the immersive viewing experience. However, videos captured by hand-held mobile cameras usually contain heavy shakiness and large parallax, which are challenging to stitch. We propose a novel approach of video stitching and stabilization for videos captured by mobile devices. The main component of our method is a unified video stitching and stabilization optimization that computes stitching and stabilization simultaneously, we can obtain the best stitching and stabilization results relative to each other without any bias to one of them. To make the optimization robust, we propose a method to identify the background of input videos. This allows us to apply our optimization in background regions only, which is the key to handle large parallax problem. We further propose a method to distinguish between right and false matches and encapsulate the false match elimination scheme and our optimization into a loop, to prevent the optimization from being affected by bad feature matches. Experiments
on a diverse of examples show that our results are much better than (challenging cases) or at least on par with (simple cases) the results of previous approaches.

Multi-view video synopsis aims at generating a brief representative video from multiple long inputs. We propose to solve the problem by joint object-shifting and camera view-switching. We synchronize the input videos and group the same object in different videos together. Then we shift the groups of objects along the time axis to obtain multiple synopsis videos with condensed activities. After that, our system automatically selects the most appropriate frame from one of the synopsis videos at each time, and all the selected frames make up our final synopsis result. We construct a simultaneous object-shifting and view-switching optimization framework to obtain the best object shifting result and view selection result with respect to each other. We also propose an alternative optimization strategy composed of graph cuts and dynamic programming to solve the unified optimization. Our experiments demonstrate the effectiveness and user convenience of our multi-video synopsis method. In the future work, we will work on more complex videos stitching to deal with large close-up moving objects, which occupy most of the video frame and interfere with the accurate estimating of camera path. We will support better user interface to control the multi-view synopsis browsing, for example, the users can generate a video synopsis with a different preference for different classes of objects.
論文摘要：
標題：魯棒多視點視頻摘要與全景導航
提交者：張鎮嵩
學位：博士
香港中文大學，二零一八年八月

隨著多視點視頻的快速增長，如何有效、高效地組織和瀏覽這些內容成為了新的研究挑戰和機遇。向用戶提供沉浸式的和高效的多視角視頻瀏覽體驗正成為一個日益重要的問題。全景顯示和視頻摘要是解決這一問題的有效方法。然而，現有的大視角全景顯示方法通常是基於計算密集型的3D重建，並且用戶難以理解同時顯示的多個視頻摘要。為了解決這些問題，本論文重點研究了用於全景顯示的街景插值和動態視頻拼接、緊湊型多視角視頻摘要的魯棒高效方法。這些方法可廣泛應用於多視點監控視頻，大視角全景顯示，以及移動相機的應用等領域。

寬基線街道視圖插值是有用的，但非常具有挑戰性。現有的方法要么依賴於重量級3D重建，要么依賴計算密集型深度網絡。我們提出了一種輕量級的高效方法，利用單應計算和優化操作來估計輸入視圖之間的分段平滑單應。我們首先對輸入圖像進行超像素分割，並基於可靠和不可靠的超像素區分進行單應性擬合和單應性傳播，該框架大大提高了估計的單應性的準確性和魯棒性。我們然後整合了單應性和網格變形的概念，提出了一種新穎的單應性約束變形公式，該公式通過利用變形網格的一階連續性來強化相鄰單應性之間的平滑性。這進一步消除了重疊，拉伸等小偽影。我們提出的方法改進了現有技術，並證明了單應性計算對於多視圖插值是足夠的。我們的城市和農村數據集的實驗驗證了我們的方法的效率和有效性。

視頻拼接可生成具有大視野的全景視頻，這可以本質上提高沉浸式觀看體驗。然而，手持式移動攝像機拍攝的視頻通常包含嚴重的抖
動和大視差，對這樣的視頻進行拼接是非常有挑戰性的。我們針對移動設備捕獲的視頻提出了一種新穎的視頻拼接和穩定方法。我們的方法主要組成部分是一個統一的視頻拼接和穩定優化，該優化可以同時計算拼接和穩定性，我們可以獲取彼此之間最好的拼接和穩定結果，而不會對其中的任何一個產生偏差。為了使優化更魯棒，我們提出了一種識別輸入視頻背景區域的方法。這使我們能將我們的優化只應用於背景區域，這是處理大視差問題的關鍵。我們進一步提出了一種區分正確匹配和錯誤匹配的方法，並將錯誤匹配消除方案和我們的優化封裝在一一個循環中，以防止優化受到不良特徵匹配的影響。對各種實例的實驗表明，與以前的算法相比，我們的結果在具有挑戰性的案例的結果要好得多，在簡單的案例的結果至少一致。

多視點視頻摘要的目的是從多個長輸入生成一個簡短的代表性視頻。我們提出聯合對象移位和像機視點切換來解決這個問題。我們對輸入視頻進行同步，並將不同視頻中的同一對象分組在一起。然後，我們沿時間軸移動這些對象組，以獲得濃縮活動的多個摘要視頻。之後，我們每次從一個摘要視頻中選擇最合適的幀來構成我們的最終摘要視頻。我們構造了一個同時進行對象移位和視點切換的統一優化框架，以獲得最佳的對象移位結果和視點選擇結果。我們還提出了一個由圖割和動態規劃組成的迭代優化策略來解決該優化問題。我們的實驗證明了我們的多視頻簡介方法的有效性和用戶便利性。在未來的工作中，我們將處理更複雜的視頻拼接，例如視頻中包含大型特徵運動對象，這些對象佔據了大部分視頻幀並干擾了攝像機路徑的精確估計。我們將支持更好的用戶界面來控制多視點視頻摘要瀏覽，例如，用戶可以針對不同類別的對象生成具有不同偏好的視頻摘要。