

3D RECONSTRUCTION

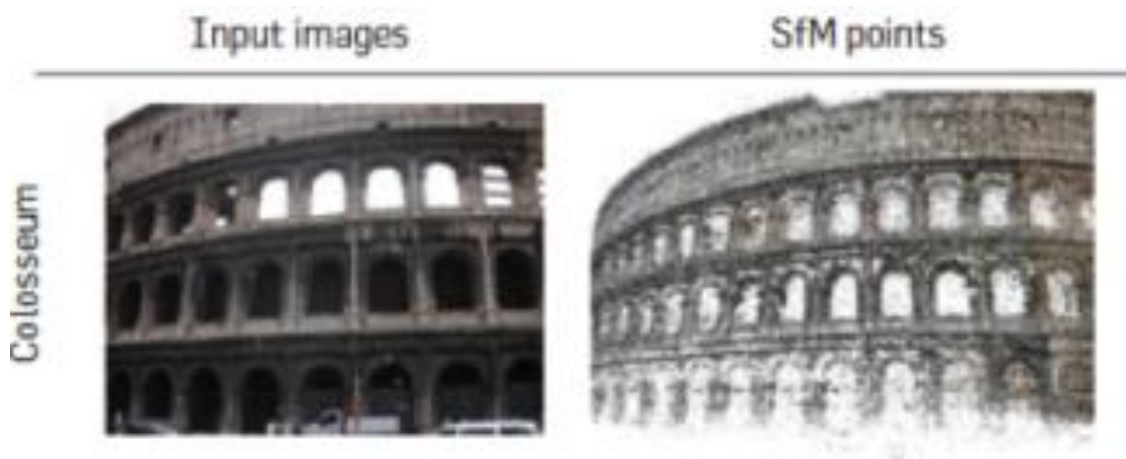
เป็นการขึ้นโมเดล 3 มิติจากรูปภาพ 2 มิติโดยอ้างอิงจาก paper ของ Shawn McCann

Ref : http://web.stanford.edu/class/cs231a/prev_projects/CS231a-FinalReport-sgmccann.pdf

จะมีการสร้างโมเดล 3 มิติ 2 แบบหลักๆด้วยกันคือ

STRUCTURE FROM MOTION

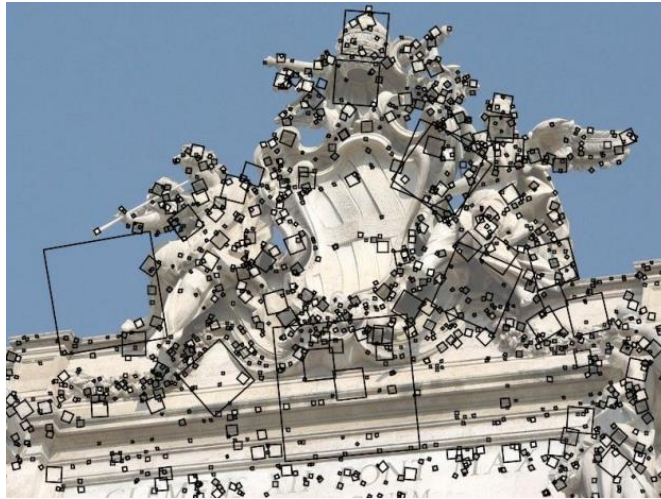
เพื่อหาขอบของโมเดล 3 มิติโดย output โมเดลที่ได้จะออกมาเป็นรูปแบบโครงโมเดลดังรูปที่ 1



รูปที่ 1 แสดงถึงผลลัพธ์จากการทำ Structure from motion

โดยจะเห็นอย่างชัดเจนว่าโมเดลที่ได้ออกมาจากรูป Input image นั้นจะออกมาเป็นโครงโดยวิธีการของ Algorithm Structure from motion คือ

1. เริ่มจากการหา Input Image โดยการขึ้นโมเดล 3 มิติโดยใช้อัลกอริทึมนี้เราไม่จำเป็นต้องทำการตั้งมุมกล้องเองจะเป็นกล้องมุมไหนก็ได้เพียงแค่ให้มีรูปของสิ่งที่เราต้องการในรูปภาพโดยใน paper นี้เขาจะนำ Input image มาจากเว็บ www.flickr.com ซึ่งเป็นศูนย์รวมรูป
2. เป็นการหา Feature ของรูปภาพ input image โดยใช้ SIFT เป็นตัวการสกัด feature ดังรูปที่ 2



รูปที่ 2 การสกัด feature โดยใช้ SIFT

3. การ match feature จะใช้ RANSAC เข้ามาช่วยทำให้การ match มีความแม่นยำมากขึ้นหลักจาก match ก็ค่อยๆขึ้นรูปโมเดลไปเรื่อยๆจนได้ผลลัพธ์ดังรูปที่ 3

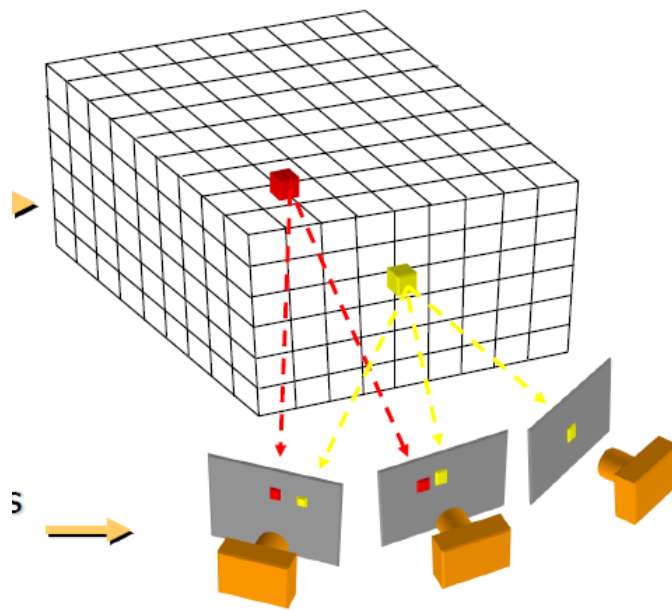


รูปที่ 3 ผลลัพธ์จากการทำ Structure from motion

ซึ่งจากรูปจะเห็นว่าโมเดล 3 มิติในอัลกอริทึมนี้ต้องใช้ Input image มหาศาลซึ่งทำให้มันช้ามากนั่นเอง

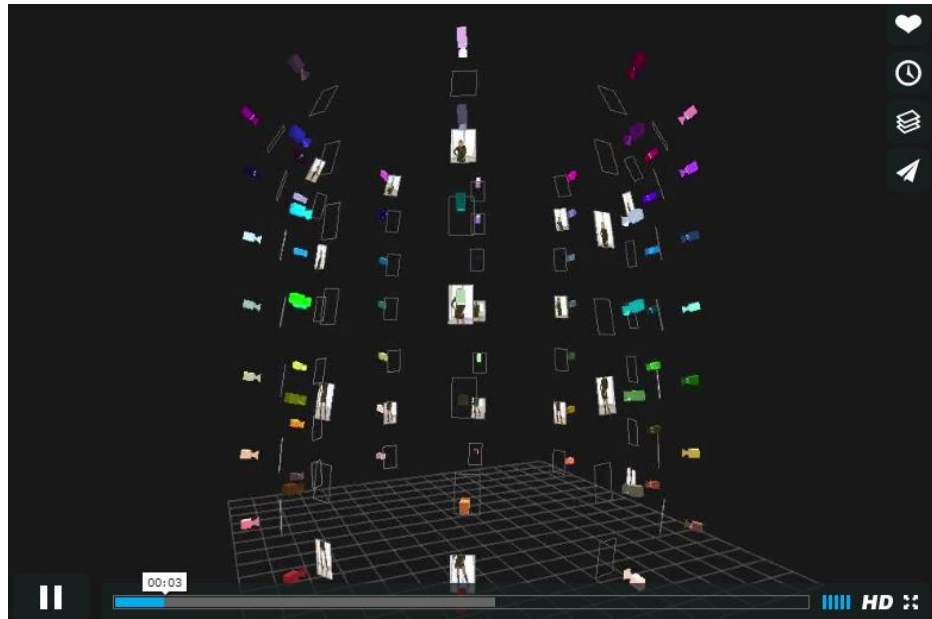
MULTI - VIEW STEREO

อัลกอริทึมนี้จะมีด้วยกันหลายแบบหลายวิธีแต่ที่ผมจะนำเสนอเป็นวิธี Volumetric stereo (Space Carving) ซึ่งเป็นแบบ Volumetric stereo แบบหนึ่งซึ่งพอมายึดถึงอัลกอริทึมดังกล่าวจะมีศัพท์หนึ่งที่ใช้กับอัลกอริทึมนี้คือ Voxel มันคือ Pixel ในรูปแบบ 3 มิตินั่นเองโดยปกติ Pixel จะถูกแทนด้วย (x, y) ดังนั้น Voxel ก็คือ Pixel ที่เพิ่มมิติเป็น (x, y, z) นั่นเอง และอีกศัพท์หนึ่งนั่นคือ Photo - consistency นั่นคือจุดใน Voxel เดียวกันที่มาจาก Input ต่างรูปกันดังรูปที่ 4



รูปที่ 4 แสดงถึงความหมายของ photo - consistency

จากรูปจะเห็นได้ชัดเจนว่าใน Input image ที่มีสีแดงจะชี้ไปยัง Voxel สีแดงโดยจะมี 2 Input image ที่ชี้ไปบน Voxel เดียวกันแสดงว่าจุด Voxel นั้นมีคุณสมบัติ photo - consistency นั่นเองโดย Input image ของวิธีนี้จำเป็นต้องมีการ calibrate camera หรือต้องตั้งกล้องเองดังรูปที่ 5



รูปที่ 5 แสดงถึงมุมมองต่างๆที่ถูกตั้งไว้

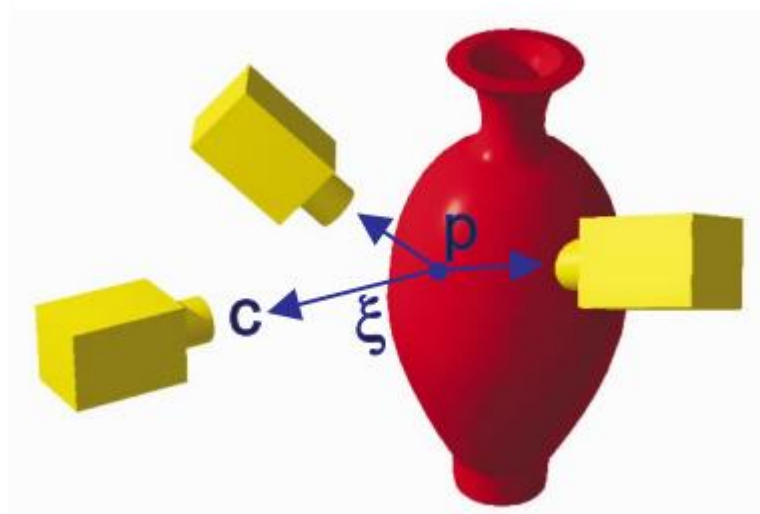
โดยถ้าดูจากมุมมองแล้วนั้นเราจะมีข้อมูล Input ด้วยกัน 3 แบบคือ

P : จุด pixel ที่อยู่บนโมเดล 3 มิติ (Voxel)

C : มุมกล้องนั้นๆ ที่ใช้เป็นตัวบอกว่ากล้องถูกมองไปทางทิศใดของตัวโมเดล 3 มิติ

ξ : คือ Vector ที่ชี้กลับไปยังมุมมอง C

ดังรูปที่ 6



รูปที่ 6 อธิบายถึง Input ต่างๆที่นำมาคำนวณ

โดยวิธีขั้นตอนการทำของเจ้าตัว Volumetric stereo (Space Carving) นั้นมีด้วยกัน 3 ขั้นตอนหลักคือ

1. Initialize volume คือสร้างรูปทรงลูกบาศก์ออกมาก่อน
2. หาจุด Voxel ที่ไม่มีคุณสมบัติของ photo - consistency โดยใช้ Input 3 ตัวมาคำนวณหาว่าจุดนี้บนรูปภาพ input image 2 มิติจะอยู่ที่ไหนบน Voxel ลูกบาศก์ที่สร้างขึ้นและถ้ามี input image อื่นมีจุด Voxel ที่ตรงกันและคุณสมบัติของรูปคล้ายๆกันก็จะทำให้จุด Voxel นั้นมีคุณสมบัติ photo - consistency นั้นเองดังนั้นในข้อนี้เราจะหาจุดที่มันไม่มีคุณสมบัติของ photo - consistency
3. เราจะนำจุด Voxel ที่ไม่มีความเป็น photo - consistency ออกไปและกลับไปทำข้อ 2 ใหม่จนกว่าจะไม่เจอจุดที่ไม่มีคุณสมบัติ photo - consistency นั้นเอง

ซึ่งทั้งสองแบบใน paper ของ Shawn McCann นั้นจะทำงานร่วมกันโดย แบบ Structure from motion จะใช้เพื่อหาขอบของโมเดล และ Multi - view stereo จะใช้เพื่อหาความหนาแน่นของโมเดลโดยเมื่อนำมารวมกันจะทำให้เกิดประสิทธิภาพที่สูงขึ้นนั่นเอง

REFERENCE

1. http://web.stanford.edu/class/cs231a/prev_projects/CS231a-FinalReport-smcann.pdf
2. http://vhosts.eecs.umich.edu/vision//teaching/EECS442_2012/lectures/Bundle_Adjustment.pdf
3. http://cs.nyu.edu/~fergus/teaching/vision/11_12_multiview.pdf
4. <http://homes.cs.washington.edu/~seitz/papers/kutu-ijcv00.pdf>
5. <https://vimeo.com/148179027>