

三、研究計畫內容（以中文或英文撰寫）：

（一）研究計畫之背景。請詳述本研究計畫所要探討或解決的問題、重要性、預期影響性及國內外有關本計畫之研究情況、重要參考文獻之評述等。如為連續性計畫應說明上年度研究進度。

1. 動機

虛擬實境（Virtual Reality；VR）是資訊技術的革命，是從大型主機、個人電腦、筆電、平板、手機等等之後的下一個裝置，是下一個資訊技術發展的進程，也是資訊科技很重要的互動點。在硬體部分，Oculus、Samsung、HTC、Sony、Google 等公司均已推出了自己的虛擬實境頭戴式顯示裝置（Head Mount Display；HMD）。在內容經營方面，虛擬實境不但可以提供一般觀眾無法體驗到的視角，像是粉絲可以透過虛擬實境的技術讓自己可以享受到像是身處搖滾區的體驗（如圖 1），虛擬實境更打破了時間跟空間的距離，移除了許多障礙與限制；社群媒體也會因為虛擬實境大幅改變，從現在的點對點的溝通，進階到包括文字、圖片、影片等所有環境的共享。根據這些應用，不僅使得全景圖片需要即時地去拍攝以及傳輸讓觀眾能享受這種身歷其境的感覺，另外，提供快速、高解析度地串接（stitching）、渲染（rendering）以及互動的方法對於越來越講求沉浸式的虛擬實境體驗來說更是極其重要。

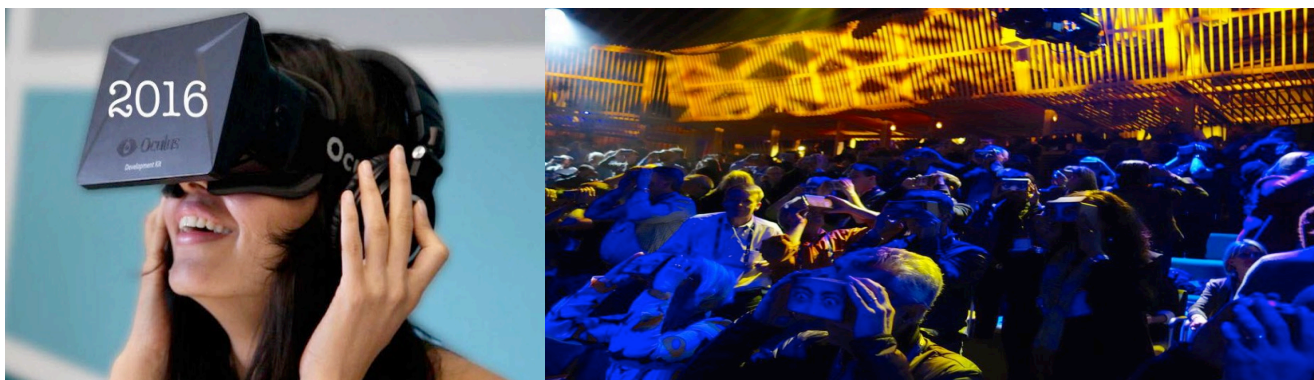


圖 1：頭戴式顯示裝置

虛擬實境的蓬勃發展也刺激了電腦圖學等相關研究的發展，過去在呈現上大多只能透過單一視角的影像與平面顯示器來展現成果，其互動性與彈性受到了視角範圍與顯示器的限制。然而透過新的全景影片以及頭戴式顯示裝置的發展，研究者開始探討如何將成果以 360 度的影片（簡稱：360 影片）做呈現。由於 360 影片不受拍攝視角所限，使用者可利用頭戴式顯示裝置與虛擬世界的人物與角色互動，也會有較好的沉浸效果，不但易於學習、富有樂趣之外，也有助於讓使用者對數位內容有更深刻的感動與瞭解。

因此，我們好奇虛擬實境的應用會為各種傳統電腦圖學與人機互動的研究領域帶來哪些挑戰與質變。傳統的影像生成與串接（capture and stitch）系統著重在如何透過減少影像間變形量（distortion）以達成生成無視差之全景圖拼貼的效果，但是串接影片以及顯示的技術仍然是非常耗時，以及在超高畫質（如：4K）下的處理速度無法提供即時串流。而在傳統的定位追蹤（tracking）的研究領域中，也曾著重於借助硬體與光學來進行各式各樣的追蹤研究，像是 HTC vive 的 Light-House 技術以及相關使用頭戴式顯示裝置追蹤全身的骨架的研究[RRCI16][CHCY15]，或是其他採取電

腦視覺的方法，即時對於場景進行構建然後進行追蹤。在這其中體感追蹤技術的優劣可謂是除了使用者所見畫面外最直接影響沉浸感的重要因素。人機互動的研究領域則是著重於如何改善使用者與電腦間的溝通與合作模式來幫助使用者獲取與操作各式資訊，像是提供自然易學的實體使用者介面（Tangible User Interface；TUI）與自然使用者介面（Natural User Interface；NUI）等用以改善虛擬實境的體驗。設計良好的 360 影片應該具有充份的提示來指引使用者找到視野外的趣味焦點。然而，多數是業餘者分享的影片，無法提供足夠觀賞者在 360 影片中找到趣味焦點的機會與提示。

在本計畫中，我們將針對 360 影片的即時生成、體感的定位與追蹤以及人機互動等三個不同面向，提出相關的技術挑戰，用以拓展電腦圖學與人機互動技術在虛擬實境領域的應用。在 360 影片生成方面，我們將運用人類眼睛視覺構造導致的聚焦變化，設計根據眼睛注視點的優化串接速度的方法「中央窩影像串接法（foveated stitching）」，作出高畫質且支援即時（real-time）顯示的 360 度全景影片系統。在即時追蹤方面，我們將運用使用者身上配戴的雙魚眼鏡頭（fisheye）建立立體視覺系統，並利用魚眼鏡頭先天的廣視角特性同時捕捉使用者全身肢體與週遭環境的深度與空間資訊，從而在無固定追蹤範圍（inside-out tracking）的前提下達成相對於世界座標系的六個自由度定位與相對於區域座標的全身骨架追蹤。在人機互動方面，我們將提出「Outside-In」這個視覺化技術以空間中子母畫面的預覽來呈現視野外趣味焦點（Region of Interest；ROI），這個子母畫面預覽的設計將允許使用者在預覽窗格中先行檢視，讓使用者在導航多重預覽全景圖時可以有較好的策略。我們所提出的三個應用範疇不但可以改善傳統電腦圖學與人機互動研究中的呈現結果，更將能有效的將這些領域的研究成果推廣到虛擬實境的應用技術之中。

2. 國內外有關本計畫之研究情況

在此研究計畫中，相關的研究領域可分成三大類：360 影片的生成與互動技術，以及體感追蹤的相關研究與應用。

A. 360 影片的生成

近年來，360 影片被視為新的影片標準，尤其是在行動裝置和虛擬實境，提供使用者更多沉浸式的視覺體驗。為了要得到高畫質的 360 影片，使用者可以使用利用雙鏡頭、雙方向感應裝置的特殊相機（如：RICOH Theta S、360FLY）來拍攝全 360 度的影像；然而，此類大部分的相機並不能提供到對虛擬實境體驗來說最重要的超高解析度（即 4K），因此，取而代之的是使用多個相機（如：GoPro）搭配環狀支架（如：Freedom360）來拍攝全景的影片。在此類的方法中，[PSZK15]使用無特殊結構性的相機群建立了產生全景影片的流程（如圖 2），他們最小化每個影像間的變形量使得高計算量的流程得以即時處理。[JG15]則建立了一個會根據影片時間、空間以及內容去做不同調整的串接方法，這個方法會對影片做區域性的修正並從其他的影片中找到相對的影格。此外，為了避免一些影片中的移動物件在傳統的處理上會在接縫處變形，[HCCC14]使用了動態接縫的方法來串接影片。

在另外一方面，許多虛擬實境的應用需要達到即時的傳輸效果，因此，使用快速串接的方法來輸出全景圖就變得非常地重要，目前也有很多針對快速影片串接的相關研究。[ERKA09]建立了一

套能在行動裝置上即時串接影片的系統，他們從不同的手機接收影片流並把這些影片同步後串接成單個影片，但是目前行動裝置並無法得到足夠使用在虛擬實境設備裝置所需的解析度（通常行動裝置所拍攝的影片解析度都小於 2K）。[CWL10]使用了動態的投影轉至把影片串接在一起，得到即時串接的結果，但是依然無法處理虛擬實境設備所使用的高解析度。而[LHC12]則使用 GPU 設計演算法來加速效能，雖然他們的方法可以串接得非常迅速，但是卻沒有包含曝光以及邊界糊化等的處理。



圖 2：[PSZK15]之研究成果

B. 體感追蹤

在運用深度資訊做即時地全身骨架追蹤的領域裡，最經典的文獻莫過於是微軟研究院[SFCS11]的研究成果，他們的演算法核心是首先將人體分為 31 個部份，各自以不同的顏色做為標籤，再透過隨機決策森林(Randomized Decision Forests; RDF)這種資料結構來快速將輸入的深度圖(Depth Map)裡屬於人體影像的像素點分配至正確的人體部位，從而推算出完整的人體骨架。RDF 裡每一個分枝節點都有一個與之對應的深度特徵值，藉由判斷某像素值經特徵函數運算後是否超過閾值(threshold)就能快速對其進行分類(如圖 3)。

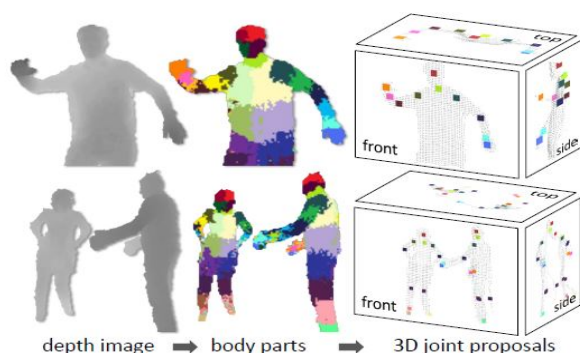


圖 3：[SFCS11]的研究成果

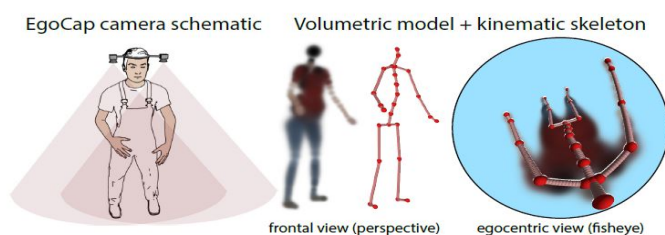


Figure 2: Schematic of EgoCap, our egocentric motion-capture rig (left), visualization of the corresponding volumetric body model and kinematic skeleton (center), and the egocentric view of both in our head-mounted fisheye cameras (right).

圖 4：[RRCI16]的研究成果

新一代的光學骨架追蹤的研究主要為針對移動式虛擬實境(Mobile VR)來增加頭部追蹤的功能，[RRCI16]使用兩台配置在頭上且用支架相連的魚眼鏡頭來建立並持續追蹤全身的骨架，其雙鏡頭

使用的 180 度視角的魚眼鏡頭能夠同時拍到使用者的全身肢體，並同時處理自身遮擋的問題。這項研究透過最佳化能量函式來推算使用者的全身骨架位置，函式包括關節角度的合理性限制、時間平滑性、基於顏色的模型對圖片相似性、深度學習網路等四項。其中主要的貢獻來自模型對圖片相似性與深度學習網路的學習，前者是藉由計算基於顏色的模型對輸入圖片的相似性來推算人體骨架，後者則是基於類似的原理引入事前訓練好的深度學習網路來輸出骨架，這部份的人體部位判別正確率可以達到 90% 左右（如圖 4）。我們在過往的研究（[CHCY15]）當中也曾利用單台魚眼鏡頭來做為人體的動作捕捉（motion capturing），因此，在這次的計畫當中，便計畫利用過往的研究經驗，來挑戰更具艱難的體感追蹤課題。

C. 360 影片的導航

在觀看 360 影片時，越來越多的研究開始探討如何建議顯示畫面的局部給使用者以獲得自然的視覺體驗。在實務上，使用者會面向螢幕或利用頭戴式顯示裝置來導航在 360 影片中的視角。在行動裝置上，[BR03]提出使用弧線來提供空間中方向和距離的指引以達成有效的視野導航。[GBGI08][GI07]也在幾年後被提出，前者針對於畫面外的移動目標做導航，而後者解決了子畫面們重疊的問題，相似概念接著被應用到擴增實境中作為畫面外實體位置的導引。另外，在視覺化（visualization）的領域特別會去探討關注的目標和畫面外目標影像在多攝影機監控環境下的最佳視野整合（如圖 5）；[GSTW07]則提出將附近影片嵌入在主影片周圍的影片播放器幫助瞭解攝影機間的空間關係，他們的研究顯示出靜態和旋轉地圖可以有效追蹤攝影機間的活動。[WKCB07]則提出一系列視覺化的方法把影片和三維空間的模型結合。



圖 5：[GSTW07]的研究成果

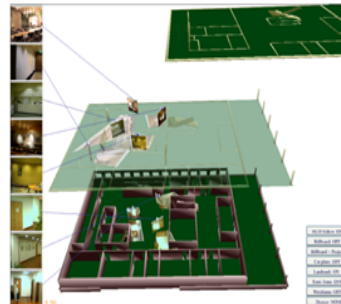


圖 6：[ALMR12]的研究成果

在另一方面，360 影片提供了周圍環境的全景視野，這在虛擬實境和擴增實境中對於提供沉浸式的遙現體驗[KR14][NYY03]和擴增人體視覺[ALMR12]上有明顯的助益。為了在廣角視野中進行更有效的搜尋，普遍應用在地圖上的「overview-plus-context」視覺化技術[TM93]來幫助使用者在探索子畫面時可以對空間資訊有更廣泛的了解。[MSDB12]則進行了一個關於長方圖、鳥瞰圖和球狀圖三種全景圖將影像中物件對應到環境的比較。[ALMR12]提升使用者在虛擬實境中的視野，並提出任何全向渲染的方法皆可以更有效的導航（如圖 6）。

3. 本計畫之目標

本計畫的主要目標便是希望能夠探索虛擬實境在三個不同面向（即影像生成、體感追蹤與互動技術）的應用，並且更進一步地能夠建構出高畫質且即時顯示的 360 度虛擬實境全景影片系統，用

以輔助使用者體驗虛擬實境時所需要的快速、高解析度串接以及渲染的要求。此外，為增加使用者與虛擬世界的互動性與資訊回饋，我們預期系統將使用者互動介面做提升，不單單是提供無纜線束縛、無固定追蹤範圍的六個自由度定位與支援即時體感追蹤，在瀏覽的過程中，輔助趣味焦點的呈現，進而讓使用者同樣地能與虛擬世界進行互動並同時提供操作虛擬世界之趣味性與方便性。因此，本計畫需要下面三個系統實作的支持：

- A. Capturing for Virtual Reality (影像生成與串接系統)
- B. Tracking for Virtual Reality (體感追蹤與定位系統)
- C. Interaction for Virtual Reality (互動導覽與顯示系統)

4. 參考資料

- ALMR12 J. Ardouin, A. Lecuyer, M. Marchal, C. Riant, and E. Marchand. FlyVIZ: a novel display device to provide humans with 360° vision by coupling Catadioptric camera with HMD. In Proceedings of ACM VRST 2012, pages 41-44, 2012.
- BCBP14 K. Buys, C. Cagniard, A. Bashkeev, and C. Pantofaru. An adaptable system for RGB-D based human body detection and pose estimation. In Journal of Visual Communication and Image Representation, 25(1), pages 39-52, 2014.
- BL07 M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. In International Journal on Computer Vision, 74(1), pages 59-73, 2007.
- BR03 P. Baudisch and R. Rosenholtz. Halo: A technique for visualizing off-screen objects. In Proceedings of ACM CHI 2003, pages 481-488, 2003.
- CHCY15 L. Chan, C.-H. Hsieh, Y.-L. Chen, S. Yang, D.-Y. Huang, R.-H. Liang, and B.-Y. Chen. Cyclops: wearable and single-piece full-body gesture input devices. In Proceedings of ACM CHI 2015, pages 3001-3009, 2015.
- CLCC09 K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu. SmartPlayer: user-centric video fast-forwarding. In Proceedings of ACM CHI 2009, pages 789-798, 2009.
- CWL10 L. Chen, X. Wang, and X. Liang. An effective video stitching method. In Proceedings of IEEE ICCD 2010, pages VI297-VI301, 2010.
- ERKA09 M. A. El-Saban, M. Refaat, A. Kaheel, and A. Abdul-Hamid. Stitching videos streamed by mobile phones in real-time. In Proceedings of ACM Multimedia 2009, pages 1009-1010, 2009.
- GBGI08 S. Gustafson, P. Baudisch, C. Gutwin, and P. Irani. Wedge: clutter-free visualization of off-screen locations. In Proceedings of ACM CHI 2008, pages 787-796, 2008.
- GFDT12 B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder. Foveated 3D graphics. In ACM Transactions on Graphics, 31(6), pages 164, 2012.
- GI07 S. G. Gustafson and P. Irani. Comparing visualizations for tracking off-screen moving targets. In Proceedings of ACM CHI 2007 EA, pages 2399-2404, 2007.
- GRU10 A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In Proceedings of ACCV 2010, pages 25-38, 2010.

- GSTW07 A. Girgensohn, F. Shipman, T. Turner, and L. Wilcox. Effects of presenting geographic context on tracking activity between cameras. In *Proceedings of ACM CHI 2007*, pages 1167-1176, 2007.
- HCCC14 K.-C. Huang, P.-Y. Chien, C.-A. Chien, H.-C. Chang, and J.-I. Guo. A 360-degree panoramic video system design. In *Proceedings of VLSI Design, Automation and Test (VLSI-DAT) 2014*, pages 1-4, 2014.
- Hirs08 H. Hirschmüller. Stereo processing by semi-global matching and mutual information. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30(2), pages 328-341, 2008.
- JG15 W. Jiang and J. Gu. Video. Stitching with spatial-temporal content-preserving warping. In *Proceedings of IEEE CVPR 2015 Workshops*, pages 42–48, 2015.
- KR14 S. Kasahara and J. Rekimoto. JackIn: Integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of AH 2014*, pages 46, 2014.
- LHC12 W.-S. Liao, T.-J. Hsieh, and Y.-L. Chang. GPU parallel computing of spherical panorama video stitching. In *Proceedings of IEEE ICPADS 2012*, pages 890-895, 2012.
- Li08 S. Li. Binocular spherical stereo. In *IEEE Transactions on Intelligent Transportation Systems*, 9(4), pages 589-600, 2008.
- MR07 C. Mei and P. Rives. Single view point omnidirectional camera calibration from planar grids. In *Proceedings of IEEE ICRA 2007*, pages 3945-3950, 2007.
- MSDB12 A. Mulloni, H. Seichter, A. Dunser, P. Baudisch, and D. Schmalstieg. 360 panoramic overviews for location-based services. In *Proceedings of ACM CHI 2012*, pages 2565-2568, 2012.
- NY03 H. Nagahara, Y. Yagi, and M. Yachida. Wide field of view head mounted display for tele-presence with an omnidirectional image sensor. In *Proceedings of IEEE CVPR 2003 Workshops*, pages 1-6. 2003
- PSZK15 F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross. Panoramic video from unstructured camera arrays. In *Computer Graphics Forum*, 32(2), pages 57-68, 2015.
- RRCI16 H. Rhodin, C. Richardt, D. Casas, E. Insafutdinov, M. Shafiei, H.-P. Seidel, B. Schiele, and C. Theobalt. EgoCap: egocentric marker-less motion capture with two fisheye cameras, *ACM Transaction on Graphics*, 35(6), pages 162, 2016.
- SFCS11 J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of IEEE CVPR 2011*, pages 1297-1304, 2011.
- SLF13 J. Schneider, T. Labe, and W. Forstner. Incremental real-time bundle adjustment for multi-camera systems with points at infinity. In *Proceedings of ISPRS 2013*, pages 355-360 2013.
- TM93 L. A. Teodosio and M. Mills. Panoramic overviews for navigating real-world scenes. In *Proceedings of ACM Multimedia 1993*, pages 359-364, 1993.
- WKCB07 Y. Wang, D. M. Krum, E. M. Coelho, and D. A. Bowman. Contextualized videos: combining videos with environment models to support situational understanding. In *IEEE Transactions on Visualization and Computer Graphics*, 13(6), pages 1568-1575, 2007.

(二) 研究方法、進行步驟及執行進度。請分年列述：1.本計畫採用之研究方法與原因。2.預計可能遭遇之困難及解決途徑。3.重要儀器之配合使用情形。4.如為須赴國外或大陸地區研究，請詳述其必要性以及預期效益等。

1. Capturing for VR

我們預計實作伺服器端與使用者端來進行高畫質（即 4K）影片的串接（stitching）以及即時串流（streaming），首先，我們將從六臺高畫質的相機（將使用 GoPro Hero4）擷取影片（每秒 30 幀），並估計相機之間的相對位置。在前處理（preprocessing）的階段，伺服器端將根據相機們的相對位置推估像素（pixel）對應的投影圖（projection map）以及糊化圖（blending map），投影圖以及糊化圖可以使系統之後在串接過程中不須經過額外的運算即可直接將輸入的影片貼到對應的全景圖（即 360 影片）上。在使用者端，使用者可運用智慧型手機搭配 Google Cardboard 來做為虛擬實境的頭戴式顯示裝置（Head-Mounted Display；HMD），而在伺服器端，我們將用其螢幕的位置來推估中央窩（foveated）區域，並發展中央窩串接（foveated stitching）演算法來加速影片串接的速度。影像串接流程處理完後，伺服器端將運用網路串流到使用者端，使用者即可以透過 Google Cardboard 與裝載在其中的智慧型手機來觀看最後的 360 影片。

A. 360 影片串接前處理

a. 產生投影圖

我們將根據相機間彼此的相對位置推估出相機在空間中的位置，包含旋轉及平移的方向，我們將運用特徵點匹配（feature matching）技術搭配手動標記來得到其結果。依據計算所得到的相機相對位置，我們將推估出各個相機之投影矩陣（projection matrix）來計算像素點對應到全景圖上的位置，並在之後的串接過程中搭配使用者選擇的投影方法（球狀或長方投影法）來將影片貼到全景圖上。由於投影矩陣並不隨時間改變，其可延用到之後剩下未處理的影格（frame），產生最後的全景影片。



圖 7：影像串接裝置與輸出全景圖

b. 產生糊化圖

為了要使兩個相機彼此間的曝光度差異減少，我們將根據不同的相機推算各自的糊化(blending)程度，而為了能加速糊化的計算過程，我們將計算一個固定的糊化矩陣(blending matrix)以減少每一幀計算所造成的系統效能負擔。影像的糊化權重(blending weight)將視使用像素距離來源影片中心的 L1 距離來決定，最後我們將把每張圖的糊化權重結合在一起並正規化(normalization)便可得到每個影片各自的糊化矩陣。

B. 全景圖投影處理

根據在前處理階段所得到的投影圖與糊化圖，我們便可以產生每幀的畫面並結合成最後的全景影片。此具體過程為針對某一時間的影格中的每個像素根據投影圖去找尋其相對應的像素，並套用糊化圖到每個像素，除此之外，我們將運用[BL07]所提出的方法來分析並計算出所需調整的對應關係來處理不同相機間的曝光度問題。我們將分別對每個像素去做串接處理，且檔案輸出大小隨像數的數目呈線性成長。

C. 傳輸與渲染(rendering)

將輸入影片在伺服器端串接完成之後，我們將得到一個 360 度視角的全景影片，接下來我們將根據 H.264 的影片編碼技術透過網路串流到使用者端，而使用者端的虛擬實境頭戴式顯示裝置在接收到資料後便可解碼還原成 360 影片。接著，我們在虛擬實境的頭戴式顯示裝置中建了一個球狀的畫布並將虛擬相機放在球心，並將先前得到的 360 影片當作材質貼到畫布上，使用者就可以在我們實作的虛擬實境的環境中達到身歷其境的體驗。

D. 中央窩以及視覺顯著區計算

a. 系統延遲的估計

理想上，我們可以使用如同[GFDT12]所提出之技術，追蹤使用者專注點來減少渲染過程所需的計算複雜度，然而，在目前的虛擬實境的應用環境下，只有極少的時間可以用來處理包括全景圖以及渲染的計算流程，此外，無線的頭戴式顯示裝置(如：Google Cardboard)通常搭配著高傳輸延遲的無線網路(即 Wi-Fi)來傳輸資料，如此一來，整體系統的延遲將使得全景圖的生成時間跟不上使用者的觀看速度，並造成原本根據上一個時刻使用者的觀看點來渲染出的高畫質區域並不會在下一個時刻被使用者看到。因此，如何處理系統延遲(system delay)以及根據使用者觀看的方向來更新畫面是建置虛擬實境環境中很重要的問題。為了解決這個問題，我們打算利用系統延遲時間來決定高畫質區所要渲染的區域大小，並估計系統在不同的階段所造成的平均延遲時間，用其得到的結果來協助渲染區域的估計。我們預計分五個階段來做分析，分別是使用者傳到伺服器的網路端、串接全景圖、伺服器傳到使用者的網路端、渲染以及顯示等五個階段。為了要使我們整體達到即時顯示，我們還需搭配硬體加速的方法來降低過高的延遲問題。

b. 中央窩串接法

目前的即時串接以及渲染技術需要一次性渲染和處理全部的點，但是 4K 解析度以及將虛擬實境用在計算上將對系統產生很大的運算負擔，並且，產生延遲將會導致系統間的傳輸不同步。為了達成支援 4K 畫質的 360 度全景影片在無線傳輸下即時顯示給使用者的目標，我們根據使用者視

線觀看的位置來降低其他未被注意之處畫面品質[GFDT12]。在使用者觀看的過程中，系統從使用者端擷取感應器讀取到的定位資料，包含偏擺、翻滾、俯仰，回傳到伺服器端，並根據這些資料系統計算出在虛擬實境的顯示裝置上使用使用者觀看的方向，進而得到中央窩的估計位置。中央窩區塊大小的設置則根據系統延遲時間而決定。我們預計使用平均延遲時間來計算中央窩區域大小，並配合視覺顯著區的估計來完成串接。而這些地方在最後的輸出成果會以高畫質的影片來做渲染和串接。

2. Tracking for VR

在此一系統中，我們預計探索如何利用穿戴式雙魚眼相機之動態計算，用以支援虛擬實境之體感追蹤與互動之部份，具體而言，我們試圖提供一個即時系統來辨識人體動作以及進行骨架（skeleton）模擬，並包含了連續的姿態追蹤與分析。為了穩定獲得視訊資料的姿勢類型，在此系統當中，使用者將配戴一個雙魚眼鏡頭的感測裝置藉以捕捉人體動作，系統將進行相機之校準（calibration）與矯正（rectification）與深度影像之計算，進而由進行骨架追蹤裡，特別強化手部追蹤的精細度與準確度。最後，我們的系統將會進行快速的姿態追蹤並且輸出六個自由度（6 Degree of Freedom；6DOF）的定位結果，此一系統便可幫助使用者與虛體世界做更緊密的互動。

A. 系統建置

我們將使用的儀器設備如圖 8 所示。雙魚眼鏡頭以及相關的計算元件預計將透過三維列印的外殼與支架組合配戴在使用者身上，然後透過無線傳輸（即 Wi-Fi）將影像資料傳輸至桌上型電腦利用具有強大平行運算能力的 GPU 來做計算，最後在實驗階段會搭配 Google Cardboard 或 HTC vive 等現行的虛擬實境顯示設備來驗證系統之效能。



圖 8：系統建置將會使用之 Raspberry Pi、魚眼鏡頭與 GPU

B. 雙魚眼鏡頭校準與矯正

在校準部分，我們將估計雙魚眼鏡頭之外在參數與內在參數。外在參數為世界座標系至左鏡頭座標系的轉換矩陣（transformation matrix），加上左鏡頭座標系至右鏡頭座標系的轉換矩陣；內在

參數是單一鏡頭座標系內影像平面中心座標與影像變形失真 (distortion) 係數。我們讓攝影機從多個不同距離與角度拍攝一個黑白相間的棋盤板子，藉以推算攝影機的外在與內在參數，這部份將依據攝影機投影模型的不同而會有演算法上的差異。在矯正部分，傳統的雙眼視覺系統透過在左右影像裡尋找相對應的立體點對來利用三角測距的原理計算深度，因此，我們將兩個光軸非平行鏡頭之影像投影到兩個光軸平行的虛擬鏡頭以減少計算量。由於魚眼鏡頭和一般鏡頭的投影模型有著較大的差別，我們估計若是直接使用傳統攝影機的透視模型來做矯正，會難以在適當的影像大小下保留原先影像中靠近邊緣的部分，從而損及了使用魚眼鏡頭來擴大週邊視野的優勢，因此，我們將選擇適合的魚眼鏡頭投影模型與影像映射函式 (mapping function) 在雙魚眼鏡頭系統的矯正步驟裡。我們預計採用[MR07]所提出的統一球狀投影模型，此一模型的特色是其參數適用於魚眼鏡頭、折反射鏡頭、甚至是普通的傳統鏡頭。在影像映射函式的部分，則採用[Li08]的研究裡使用的經緯度映射法，如此一來雙魚眼鏡頭系統的影像即使做過矯正，仍然可以在適當的影像大小下保留所有資訊。

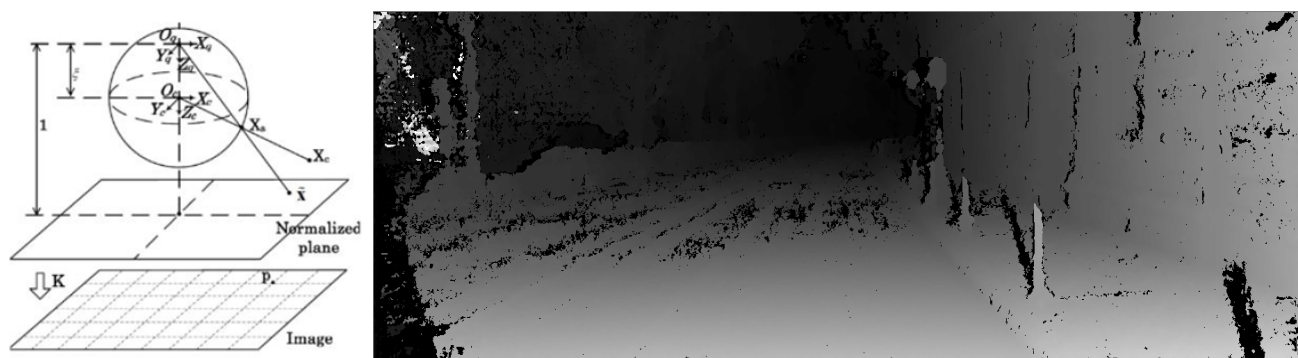


圖 9：統一球狀投影模型[MR07]與視差圖[Hirs08]

C. 動態骨架追蹤

在上一階段完成了左右眼影像的校準與矯正之後，在此一階段中，首先，我們預計採用[Hirs08]的深度估計演算法來計算出左右眼之視差圖 (disparity map) (如圖 9 右)，這裡我們將同時運用 GPU 的平行計算來大幅提升尋找立體點像素的三維空間位置。有了視差圖以後，我們將根據先前的投影模型與映射函式反向推算出視差圖中每一像素的三維空間位置，從而將視差圖轉換成三維點雲 (3D point cloud) 或深度圖。接下來，我們預計使用[BCBP14]所提出的方法進行人體骨架的追蹤，其方法同樣使用 RDF 的架構將影像中屬於人體的像素做部位分類從而推算出骨架，但透過在第一階段以整張深度圖做輸入、在第二階段以整張彩色 (RGB) 影像做輸入，來減少事先訓練 (training) 的需要。接著，我們考慮動作上的時間相依性 (temporal coherence) 修正並輸出使用者當前動作，即可以推算出連續的骨架圖。最後，我們將採用[SLF13]的方法，運用雙魚眼鏡頭影像背景的資訊做場景的三維模型重建與定位，以推得使用者相對於世界座標軸的六個自由度定位。

3. Interaction for VR

在這一個階段，我們將探索如何以視覺化 (Visualization) 技術來增強使用者對於 360 影片之瀏覽 (browse) 體驗。360 影片完整地紀錄了環境的所有內容，然而為了提供自然的觀看體驗，不

論是透過螢幕或是頭戴式顯示裝置，使用者通常只能看到其中的一小部分。當影片的趣味焦點（Region of Interests；ROI）落在可視範圍（Field of View；FOV）之外時，便會造成使用者搜尋的困難，另外使用者也可能花上許多時間在尋找不存在的趣味焦點。因此，我們預計提出一個視覺化的方法「Outside-In」，透過具空間引導的子畫面（Picture in Picture；PIP）將視窗外的趣味焦點重新引入主視窗內，並利用子畫面視窗的幾何性質進一步呈現趣味焦點與主畫面的相對方位，讓使用者可以更有效的瀏覽。為了減緩多個子畫面彼此遮擋的問題，我們使用快速序列視覺呈現的方法，確保每個子畫面在分到的時間內可以完整的被看到。我們預計透過使用者實驗找到在維持可讀性的前提下，適當的子畫面的幾何變化以及適當的顯示頻率。在實際應用上，我們也將提供了觸控螢幕上的互動方式以及遠端視訊的使用情境。

我們預計實作一個讓視野外趣味焦點顯示在視野中的視覺化技術，以提供使用者有效的預覽介面。我們將設計子母畫面不同的幾何性質來呈現趣味焦點與使用者的相對位置關係，並可分成三種參數設計。

A. 子畫面的位置

其位置只要指出視野外趣味焦點相對於使用者視角的方向，同時，為了降低阻擋到主要內容的機會，我們讓子畫面預覽位置盡量靠近邊界。首先，我們將主螢幕上的子畫面依趣味焦點的相對位置給予方向資訊。傳統應用在飛行的大圓航線（球面上兩點最短路徑）被證明是最節省燃料的路徑，因此我們將嘗試將子畫面放在視野中心和趣味焦點中心的大圓航線上，然而，因為大圓航線有時會通過極區。這的確是最短路徑，但對於人類而言脖子在這種垂直方向上的轉動是困難的（如圖 10(a)中的藍色路徑），人們會較習慣於水平方向的旋轉。

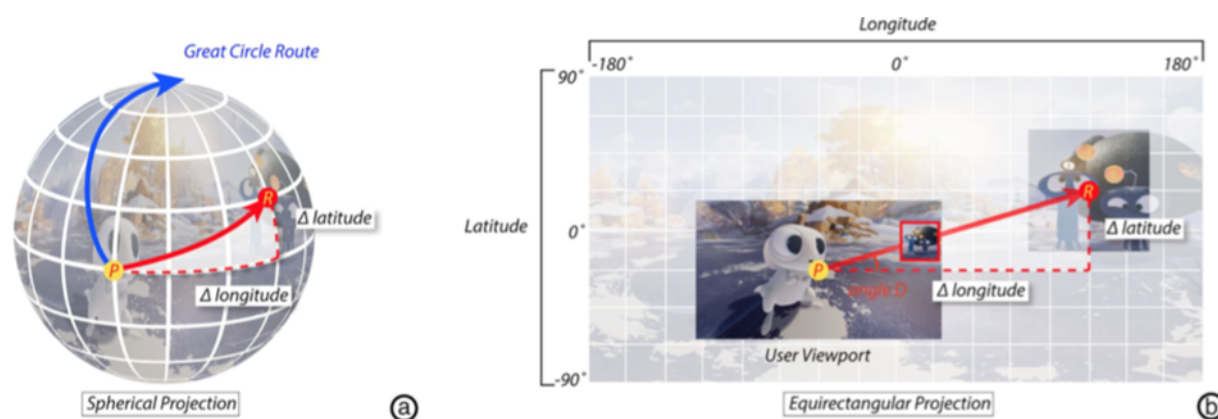


圖 10：子畫面方向的計算方式

因此，我們預計將球狀圖展開到平面上也就是等距長方投影，來重新計算方向角（圖 10(a)中的紅色三角形）。這種等距長方投影圖，經線是等距鉛直線，緯線是等距水平線。假設使用者正在看點 P（觀察點經度，觀察點緯度），視野外趣味焦點在點 R（趣味焦點經度，趣味焦點緯度），如圖 10(b)所示，觀察點和趣味焦點間的連線與 x 軸和 y 軸間的夾角組成一個直角三角形，三角形的角 D 作為方向角並將子畫面放在此連線上。

這個位置預計可以更加直覺的引導使用者的移動，原因是等距長方投影符合人體脖子轉動範圍，等距長方投影在鉛直方向的邊界為 -90 度 \sim 90 度跟人體脖子鉛直移動範圍相同，在水平方向上也跟人們脖子所能轉動的角度相似。一個極端的例子是趣味焦點位於 $(179, 89)$ 、觀察點位於 $(0, 0)$ ，大圓路徑會是通過北極的連線，這會使脖子做 91 度的轉動而使人感到不適。但是用我們的方法，三角形的斜邊會始於一個適當的角度，當觀察點越接近趣味焦點才會漸漸變陡。

獲得方向角後，我們會將子畫面盡可能的放置於接近邊界的地方以減小對主畫面的影響。為達到此效果，我們將子畫面放在長軸為螢幕寬短軸為螢幕高的橢圓軌道上（圖 11 (a)）。

B. 子畫面的旋轉

而在子畫面的旋轉方面，我們將採用兩個方向的旋轉，子畫面窗格沿著外緣的傾斜角表示和趣味焦點間的距離，這個透過傾斜角度改變的視覺提示，可以幫助使用者建立趣味焦點相對位置的粗略概念並在使用者遠離或接近時獲得回饋。另一個旋轉方向則是圍繞視野中心，讓子畫面的外緣一直面向趣味焦點。為加強方向上的提示，我們將子畫面在二維空間中依螢幕中心做旋轉。如圖 11 (b)所示，子畫面的內邊會面向視野中心，而外邊會面向趣味焦點。此外，得到的外邊會作為距離引導的傾斜軸，此將於以子畫面的旋轉作為與趣味焦點間距離的的引導說明。此外，我們將子畫面中的影像維持水平以便觀看。當子畫面依視野中心旋轉，我們將同時旋轉負責子畫面的虛擬攝影機，以讓子畫面中的影像即使框架旋轉仍維持不變。

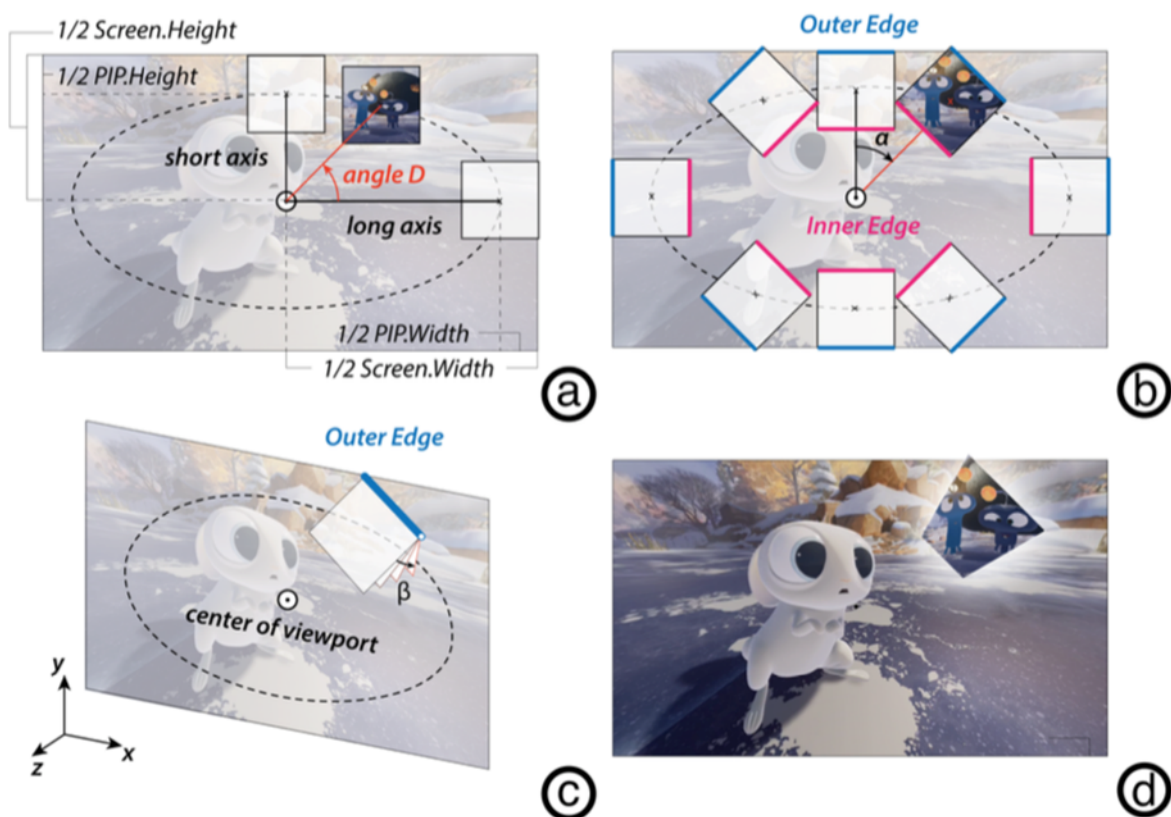


圖 11：子畫面的旋轉

顯示趣味焦點的距離，也就是使用者需要轉多少角度以找到趣味焦點，是空間導引的另一個重要功能。我們設計的主要概念藉由上一個設計所獲得的外邊來傾斜子畫面以視覺化與趣味焦點間的距離，也就是圖 11(c)的藍色軸。趣味焦點越遠其對應的子畫面傾斜角度越大。以下敘述我們如何依據趣味焦點的距離來計算傾斜角度。首先我們定義趣味焦點和觀察點的距離為等距長方投影上的歐幾里德距離，從圖 10 所示投影上的三角形，我們採用斜邊的長度作為距離，一旦我們獲得距離，即可將距離線性對應到傾斜角度。值得注意的是，子畫面傾斜越多，影像扭曲就越大，因此最大的傾斜角度需設定在合理範圍內以確保所有內容是可讀的。

C. 子畫面的不透明程度

子畫面的不透明度依趣味焦點的出現與消失來調整，為減輕重疊的問題我們用了 Time-Carousel 的技術讓子畫面輪流顯示，並讓不是目前的子畫面呈現半透明。此外，子畫面可以移動並可能突然出現或消失，一旦趣味焦點的中心離開視野中，子畫面就會出現，或者，趣味焦點的中心進到視野中，子畫面就會消失，也就是變成透明。當趣味焦點出現在視野外區域，同時子畫面就會出現，當趣味焦點消失在視野外區域，子畫面就會消失。我們也應用 Time-Carousel 技術在不透明度上以解決互相重疊的問題。另外，當有數個子畫面出現時，互相重疊是無可避免的，為減輕這個問題，我們採用 Time-Carousel 讓重疊的子畫面群輪流顯示。當不是這個子畫面時我們將其不透明度降低到百分之二十五。這樣做有兩個原因：第一，完全透明可能讓使用者混淆成趣味焦點消失或進到主視窗；第二，我們試著讓使用者知道那裡仍有趣味焦點。我們將子畫面中心投影到畫面上並計算其距離來偵測是否重疊，假如有一堆點，我們會讓他輪流顯示並使其他群體中的子畫面變淡。

(三) 預期完成之工作項目及成果。請分年列述：1.預期完成之工作項目。2.對於參與之工作人員，預期可獲之訓練。3.預期完成之研究成果（如期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4.學術研究、國家發展及其他應用方面預期之貢獻。

1. 預期完成之工作項目

A. 第一年預計完成目標

- a. 完成視覺顯著區追蹤演算法，並驗證計算出之區域為使用者視覺容易停留之處
- b. 實驗的中央窩區域跟視覺顯著區做結合，找尋出最後全景影片的高畫質區域
- c. 使用者視覺中央窩區域改成網路傳遞動態資訊給伺服器做串接

B. 第二年預計完成目標

- a. 用雙魚眼視覺系統達成「Inside-Out」配置的六個自由度定位
- b. 用雙魚眼視覺系統達成「Inside-Out」配置的基本全身骨架追蹤
- c. 視覺化焦點的介面設計，手機觸控相關應用與頭戴式顯示器實作

C. 第三年預計完成目標

- a. 使用者實驗（User Study）
- b. 整體效能分析與評估
- c. 強化骨架追蹤的精細度與準確度並提升六個自由度定位的速度

2. 對於學術研究、國家發展及其他應用方面預期之貢獻

虛擬實境的發展在近年來已經是顯而易見的趨勢，如何對一連續動作做有效的認知更是邁向人性化使用者介面的重要議題，為呼應目前的發展趨勢，本計畫在即時串拼、體感追蹤、影片視覺化等皆有所涉及與突破。任何一個單一研究皆為該領域的優良成果，並同時計畫將這三方面的研究成果拓展到虛擬實境的應用層面上。

3. 對於參與之工作人員，預期可獲之訓練

- A. 掌握電腦視覺影像串接與特徵點追蹤之相關技術。
- B. 掌握虛擬實境的使用者經驗研究的相關技術。
- C. 掌握人機互動介面領域與全景影片視覺化技術的相關技術。
- D. 獲得大型軟體函式庫與系統發展的實作經驗。
- E. 對於相關的基礎知識，如三維投影模型、相機參數校正與平行運算等有更深入的了解。

(四) 整合型研究計畫說明。如為整合型研究計畫請就以上各點分別說明與其他子計畫之相關性。

本計畫非整合型研究計畫