

TapID: Rapid Touch Interaction in Virtual Reality using Wearable

Sensing

Yi-Xiang Qin (覃毅翔)

410985006

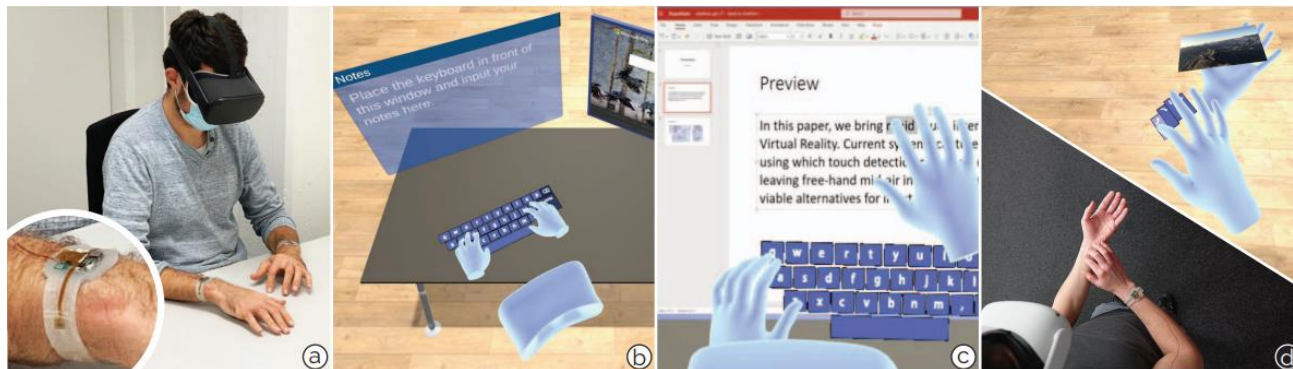


圖 1: TapID 是一種戴在手腕的設備，可以識別手指在平面上的敲擊，配合手部姿勢追蹤，可以轉換為 VR 中的輸入。(a)用戶佩戴兩個 TapID 手環(b)與 VR 中的平面物件進行互動，例如用於文本輸入、網頁瀏覽(c)或使用前端應用程式編寫文檔。(d)身體本身也可以被視為平面，使用 TapID 來檢測手指在身體上的敲擊，範例中可以旋轉手中的圖像。

概述

當前的虛擬現實系統通常使用相機來捕捉控制器的輸入或徒手在空中的運動。這是讓 VR 作為生產力工具的關鍵障礙，因為控制器或空中徒手輸入容易導致疲累，無法長時間持續操作。既然用戶已經習慣於在手機和平板電腦上使用的輸入模式，就將表面上的觸控引入虛擬現實。因此我們提出了 TapID，這是一種基於手腕的慣性傳感系統，它補足了以往 VR 眼鏡依靠手部姿勢觸發 VR 輸入的缺點。TapID 可以準確地檢測平面觸控事件，並識別用於觸摸的手指。此篇報告將會從為什麼需要在 VR 中使用空手操作的原因開始談起，接著描述達成此目標的技術原理及需要用到的物理設備，最後說明其最新的技術發展及相關應用。

1. 問題描述

以前的設備高度依賴於手持控制器的輸入，但在商業應用等領域上，操作方式向無控制器、手勢和手勢操作輸入的轉變是顯而易見的。在擴增實境中，使用手部跟蹤技術時可以讓用戶抓住物體，從而防止使用過程中的疲勞。但在虛擬實境中，雙手在空中大量操作時容易感到疲累，所以將虛擬實境中的所有直接互動轉移到與平面之間的互動，可以有效降低長時間操作時的疲勞感。同時可藉由平面提供的觸覺反饋，快速和精確的判斷操作。

而虛擬實境中的手是由頭戴式裝置本身進行追蹤的，單一位置的相機導致測得的深度不精準，本文通過用戶佩戴在手腕上的 TapID 手環來協助判斷快速的觸控操作。TapID 可以通過配置的慣性傳感器，來檢測平面上的觸摸事件，並辨認出觸控的手指。

利用 TapID 偵測到的觸摸事件，結合頭戴式裝置針對手、手指的光學追蹤，可以觸發虛擬實境裝置中的輸入事件。因此，我們可以將平板電腦、手機上的應用轉移到虛擬實境的廠經中使用。

2. 方法

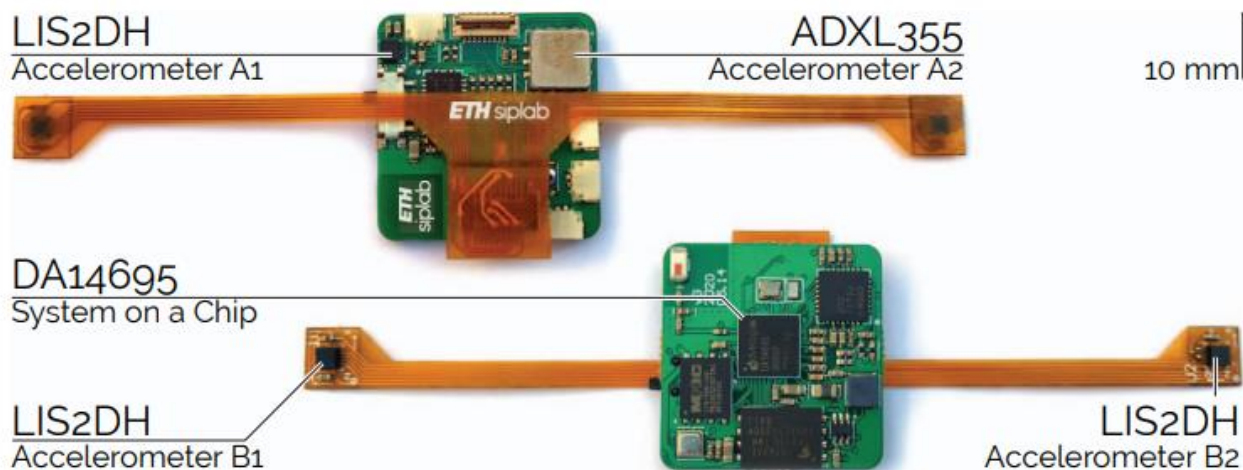


圖 2：TapID 電子平台。主 PCB 具有一個微控制器，可讀取安裝在柔性 PCB 末端的兩個低功耗 IMU。為了在我們的評估中進行比較，在主 PCB 上安裝了兩個額外的 IMU。

圖 2 顯示了 TapID 電子平台，它以系統單晶片（DA14695，Dialog Semi）為中心，以 1344 Hz 的頻率對所有慣性傳感器進行採樣，並將 IMU 數據流傳輸到 PC。TapID 具有兩個通過柔性 PCB 嵌入腕帶的低功耗加速度計（LIS2DH，STMicroelectronics）以及兩個裝在板上用於測試和比較不同傳感器配置的加速度計。使用 Shore-32 矽膠澆鑄 TapID 的錶帶，並在固化過程中嵌入所有電子設備。

每次當 TapID 偵測到觸控事件時，系統會將處理好的時間戳記手指的及世界座標表示的 2D 位置傳給虛擬實境應用程式，此過程的總延遲大約為 130 毫秒。處理觸控事件分三個部分。分別為：

- 一、用於輕敲檢測的信號處理。
- 二、基於機器學習的識別流程。
- 三、結合 VR 頭戴式裝置與手部位置追蹤。

在第一步驟中，我們透過不同的運動在加速度計信號中，會產生不同的尖銳尖峰區域，來區分點擊事件與其他手臂和手部運動。TapID 不單純依靠大小作二值化(Thresholding)處理，而是利用指數下降(exponential decrease)的方式，只關注信號的快速變化。TapID 會累積所有加速度計軸的原始信號(Raw Signals, RS)的變化率來計算變化率分數(Rate-of-Change Score, RCS)。

$$RCS_t = \frac{RCS_{t-1}}{1.6} + \sum_{\text{sensors}} \sum_{\text{axes}} |RS_t - RS_{t-1}| \quad (1)$$

將觸控事件的時間戳記定為超過閾值後 20 個樣本(15 毫秒)中，取最高的 RCS，並在偵測到事件後會有 200 毫秒的回退期(back-off period)。這樣可以確保單次觸碰只會觸發一個事件，並在雙手同時輸入的情況下，可以允許每分鐘 600 個事件。

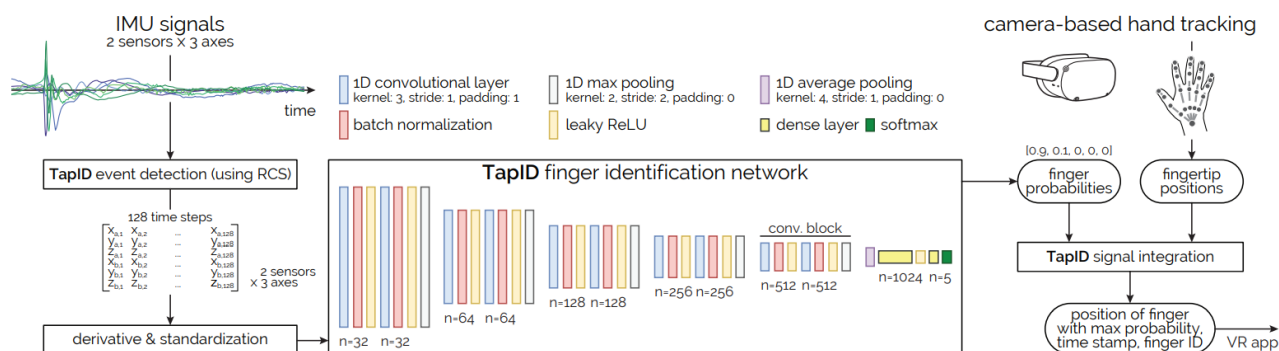


圖 3:TapID 的處理流程。我們的主要輸入是來自兩個 3D 加速度計的連續，TapID 首先使用 RCS 從中檢測到輕擊的發生。檢測後，我們提取一個包含 128 個樣本的窗口，以敲擊事件為中心，並將其輸入到我們的神經網絡分類器中以識別敲擊手指。我們的架構包括五個卷積層塊和兩個最終線性層，它們輸入到 softmax 激活函數中，從而呈現手指上的概率分佈。然後，我們將分類與輔助輸入相結合：連續手部跟踪。在確定是哪個手指引起了事件後，我們檢索其指尖位置並將點擊事件轉發給 VR 應用程序。

第二步驟會將第一步驟取得的时间戳記、加速度計的原始資料輸入置神經網路分類器來區分手指。此分類器主要以監督式學習訓練，利用反向傳播算法(backpropagation)來減少訓練集上的交叉熵(cross-entropy)損失。訓練出來的分類器可以將每個事件依不同的手指分類到五個類別。

圖三顯示了分類器的架構，主要為遵循 VGG 架構[1]的多層前饋一維卷積神經網絡，並採用更深的網路與更小的核心，由於層與層之間存在更多的非線性變化，使得神經網路可以提取更強大的特徵。此網路由五個區塊組成，每塊有兩個核心大小為 3 的卷積層，最後再連接到一個最大的池化層(pooling layer)用來減少一半的輸入維度。卷積塊後面是一個自適應的平均池化層和兩個線性層，最後的線性層將結果送入 softmax 的激活函數，函數會估計出五根手指的概率分佈。

利用左右手的對稱性，交換左手 TapID 中的兩個加速度計的數據，並根據傳感器的方向反轉 x 或 y 軸，可以接收來自雙手的輸入信號。這減少了組合數據集的變動程度並有助於分類器的學習過程，訓練完的分類器可以應用於來自任一個 TapID 的信號。

第三步驟利用市面上有手部追蹤功能的頭戴式裝置(Oculus Quest)，VR 系統使用 Unity 的 3D 遊戲引擎進行渲染，並使用 Velt[2]提供所有數據。在第二階段確定了哪隻手指觸發了點擊事件後，會從一組被跟蹤的手中檢索指尖位置，並將其分配到 TapID 的點擊事件中。若檢測到的指尖不在 VR 中的表面附近，系統會將其視為無效輸入。

3. 發展



圖 4: TapID 與 VR 場景的結合用於操作不同的平面小物件。

TapID 與 VR 介面結合可以達成許多輸入指令的操作，用戶可以通過用特定手指點擊表面上所需的位置來打開和定位物件。例如，左小指可以叫出方向鍵，而右小指觸發數字鍵盤出現，拇指可以在雙手之間召喚一個簡單的鍵盤。

照片排序(圖四 a):透過將調整圖片大小、旋轉圖像、移動等指令分配給每個手指。此程式在沒有桌子的情況下也可以透過將物件錨定在另一隻手上來完成(見圖一 d)。

文檔編輯(圖四 b):可以透過虛擬鍵盤打字、雙擊選擇字或是句子來更改格式，利用左小指叫出的方向鍵也可以快速調整文本游標的位置。

數字輸入(圖四 c):可以利用右小指快速地開啟及關閉數字鍵盤，進行數字輸入。

鋼琴彈奏(圖四 d):充分利用 TapID 進行準確的輸入，讓用戶可以用任何手指敲擊鋼琴鍵。VR 中的視覺、聽覺反饋、平面的觸覺反饋以及 TapID 方法的低延遲相結合，能夠在演奏這種虛擬樂器的同時保留用戶敲擊輸入的準確節拍。

引用

- [1] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv 2014
- [2] Andreas Fender and Jörg Müller, “Velt: A Framework for Multi RGB-D Camera Systems”, association for Computing Machinery 2018