

Frei form-Gesten

Abstract

In diesem Dokument wurden Erfahrungen über die Definition von Frei form Gesten aus dem MMI Projekt „Earth Explorer“ und der Weiterentwicklung zusammengefasst. Die Erfahrungen im MMI Projekt kommen eher aus der praktischen Richtung. Die Weiterentwicklung ist noch nicht funktionstüchtig implementiert und ist daher eine Zusammenfassung einer Internet Recherche.

Definieren von Gesten

Definieren der Gesten kann überraschend schwierig sein. Die einfachsten Gesten sind diejenigen, die zu einem einzigen Zeitpunkt geschehen, und daher nicht auf vergangene Positionen der Gelenke zurückgreifen müssen. Zum Beispiel, wenn der Benutzer seine Hand über seinen Kopf hebt. Dies kann in einem einzelnen Frame überprüft werden. Komplizierte Gesten brauchen eine Zeit um als solche erkannt zu werden. Für eine zoom Geste, ist es nicht möglich aus einem einzigen Frame zu erkennen, ob eine Person zoomt oder einfach nur die Hand vor den Körper hält.

Das Programm muss in der Lage, relevante Informationen aus der Vergangenheit zu speichern, nur welche Informationen sind relevant? Sind die letzten 30 Frames genug Information? 30 Bilder bekommt man in nur einer Sekunde, vielleicht 60 Frames, also 2 Sekunden? Oder 5 Sekunden, also 300 Bilder? Menschen bewegen sich nicht so schnell. Vielleicht könnte man jeden fünften Frame speichern, dadurch würde die Länge bei 5 Sekunden wieder auf 60 Frames reduziert werden.

Eine bessere Idee wäre die relevanten Informationen aus den Frames zu wählen. Für eine zoom Geste der Hände, die aktuelle Geschwindigkeit, wie lange sich bewegt wurde, wie weit wurden die Hände verschoben, etc. könnten nützliche Informationen sein.

Nachdem das definiert wurde, speichert man alle Informationen die im Zusammenhang mit der Geste stehen. Doch wie definiert man diese Informationen in Zahlen? Zoomen könnte eine bestimmte Mindestgeschwindigkeit, oder eine Richtung (links / rechts nach außen), oder eine Dauer erfordern.

Hier sind jedoch nicht die definierten 5 Sekunden von Interesse. Diese Dauer beschreibt das absolute Minimum das benötigt wird um davon auszugehen, dass der Benutzer zoomt. Wie bereits erwähnt, kann das nicht anhand eines einzelnen Frames bestimmt werden. Es sollte nicht allein die Implementierte Dauer bestimmen, ob eine zoom Geste als solche erkannt wird. Da die Zeitspanne in der die Geste durchgeführt wird für jeden Benutzer unterschiedlich ist, da jeder Benutzer sich unterschiedlich schnell bewegt.

Die Chancen stehen gut, dass an einem gewissen Punkt innerhalb dieser 5 Sekunden die Geste mit der Definition übereinstimmt und ein zoomen erkannt wird. Es gibt aber ein weiteres Problem, wie definiert man den Start und Endpunkt einer Geste? Die Geste kann während des Ausführens langsamer und schneller werden oder gar pausieren. Fängt die Geste nach der Pause wieder von Anfang an?

Der Punkt, den ich versuche hier klar zu machen ist, es gibt keine einfache Möglichkeit Gesten zu erkennen. Als Entwickler muss man sich in die Geste hinein denken und eine Reihe von gemeinsamen Positionen der Skelett-Daten über eine Zeit hinweg als eine Geste zu definieren.

Benutzererwartung

Der Benutzer sollten nicht gezwungen werden eine Bewegung für einen bestimmten Zeitraum zu wiederholen. Es ist überraschend anstrengend und nicht einfach. Der Benutzer erwartet eher wie beim Computer eine Art „Point-and-Click“, sobald geklickt wird, wird eine Reaktion des Systems erwartet. Es ist sehr unnatürlich einen Mausklick (z.B. 5sekunden) halten zu müssen bevor eine Anwendung startet.

Wiederholen einer Geste über einen Zeitraum ist okay, wenn es sich um eine stetige Aktion handelt und stetig Rückmeldung vom System gegeben wird. Somit versteht der Benutzer, dass für diese Aktion eine kontinuierliche Ausführung erforderlich ist.