

Adaptive Gestenerkennung mit Variationsabschätzung für interaktive Systeme

Maxim Boianetchii¹ und Marian Stein²

¹ Universität Rostock
`maxim.boianetchii@uni-rostock.de`

² Universität Rostock
`marian.stein@uni-rostock.de`

- 1 Einführung**
- 2 Ähnliche Arbeiten**
- 3 Interaktionsprinzipien**
- 4 Zugrunde liegendes Modell**
- 5 Erkennung von realen 2D-Gesten**

Um die Gestenauswertung zu evaluieren, wurde die Gestendatenbank von Wobbrock et al.[1] verwendet. Diese enthält Daten von 16 Stiftgesten, die von zehn Teilnehmern in drei verschiedenen Geschwindigkeiten jeweils zehn mal aufgenommen wurden. Für die Versuche wurden pro Geste zufällig aus den Daten eines Teilnehmers jeweils ein Trainings- und ein anderer Testdatensatz ausgewählt. Insgesamt wurden 4 verschiedene Tests durchgeführt, die jeweils 100 Mal wiederholt wurden:

1. Gleiche Testbedingungen, wie sie von Wobbrock[1] verwendet wurden.
2. Einfluss von geänderten Verteilungsparametern.
3. Verwendung von Trainings- und Testdaten mit unterschiedlichen Geschwindigkeiten.
4. Genauigkeit der Erkennung von Gesten, während sie noch nicht abgeschlossen sind.

Bei allen Tests außer 3. wurden Trainings- und Testdaten mit der gleichen Geschwindigkeit gewählt. Da GVF auch Veränderungen in den Gesten erkennen und ausgeben soll, ergibt sich für diese Tests ein Zustandsraum x_k , der aus der Phase p_k , der Geschwindigkeit v_k , der Skalierung s_k und dem Drehwinkel α_k besteht.

5.1 Erkennung von Beispielen gleicher Geschwindigkeit

Im ersten Test wurde die durchschnittliche Erkennungsrate sowie die Standardabweichung von GVF mit den erreichten Werten von \$1 Recognizer[1], DTW und GF verglichen. Im Vergleich zu GF erreicht GVF eine bessere Erkennungsrate (98,11% gegenüber 95,78%), was sich darauf zurückführen lässt, dass GVF sich an Skalierung und Drehung anpasst. Auch im Vergleich mit den beiden Offline-Methoden erreicht GVF leicht bessere Erkennungsraten (98,11% gegenüber 97,27% bzw. 97,86%).

5.2 Einfluss der Verteilungsparameter auf die Erkennung

Im zweiten Test wurde der Einfluss der Standardabweichung σ und der Parameter v der Student'schen t-Verteilung auf die Erkennungsrate untersucht. Untersucht wurden dabei für σ in Zehnerschritten Werte von 10 bis 150 und für v die Werte 0.5, 1, 1.5 und ∞ (hier entspricht die Verteilung einer Gaussverteilung).

Die Ergebnisse sind in Graph X zusammen mit den Erkennungsraten der beiden Offline-Methoden dargestellt.

Hier sind zwei Beobachtungen festzustellen. Erstens, dass die Erkennungsrate die Beste ist für beschränkte σ und v . Zweitens, dass die Verwendung einer Student'schen t-Verteilung anstatt einer Gaussverteilung den Vorteil hat, dass die Erkennungsrate wesentlich weniger von σ abhängt.

5.3 Erkennung von Beispielen verschiedener Geschwindigkeit

Im dritten Test wurden die Erkennungsraten von GVF mit denen des \$1 Recognizers verglichen. Diesmal wurden allerdings die Vorlage und die Testgeste aus verschiedenen Geschwindigkeitsgruppen verwendet. Tabelle Y zeigt die Erkennungsraten der unterschiedlichen Kombinationsmöglichkeiten. Es wird deutlich, dass insgesamt die durchschnittliche Erkennungsrate beider Algorithmen vergleichbar ist (94,6% gegenüber 94,8%). Ebenso fällt auf, dass die schlechtesten Ergebnisse erreicht werden, wenn Vorlage und Testgeste entgegengesetzte Geschwindigkeiten haben, mit Genauigkeiten von 88,3% bzw. 85,9%.

5.4 Früherkennung der Gesten

Zuletzt wurde die Erkennung der Gesten untersucht, während sie ausgeführt werden. In Graph Z sind die Früherkennungsraten von GVF und GF in Abhängigkeit vom Fortschritt der Geste dargestellt. Es wird deutlich, dass GVF über den ganzen Messbereich bessere Erkennungsraten erreicht als GF. So erzielt GVF im Schnitt bereits nach 10% der Geste eine Genauigkeit 67% und erreicht bereits nach 40% Fortschritt 90% Genauigkeit.

5.5 Ergebnisse der Versuche mit realen Daten

In den Versuchen hat sich gezeigt, dass GVF in der Erkennung von Gesten gleich gut bis besser ist als andere aktuelle Erkennungsmethoden. Darüber hinaus wurde deutlich, dass die im GVF verwendete Student'sche t-Verteilung gegenüber der im GF verwendeten Gaussverteilung weniger auf eine gute Abschätzung von σ angewiesen ist. Somit ist diese Methode besser in Anwendungsfällen verwendbar, in denen wenig Trainingsdaten zur Verfügung stehen, und somit σ schlecht abgeschätzt werden kann. Eine Verbesserung gegenüber bisher üblichen Methoden stellt der GVF auch dadurch dar, dass bereits während der Gestenausführung Ergebnisse geliefert werden, und die Früherkennung gut genug funktioniert, dass schon nach 10% der Geste Erkennungsgenauigkeiten von über 60% erreicht werden.

6 Abschätzung der Varianz anhand synthetischer Daten

Um die Anpassung an Varianz der Gesten auszuwerten, wurden synthetische Daten verwendet, da menschliche Eingaben keine exakten Werte für die Varianzparameter liefern würden, mit denen die Ergebnisse des GVF verglichen werden könnten. Es wurden zwei Fälle betrachtet: Im ersten Fall wurde nur die Phase und Skalierung der Gesten variiert, im zweiten Fall zusätzlich auch die Rotation. Für die Generierung der Daten wurde eine Viviani-Kurve verwendet:

$$C(t) = \begin{cases} x(t) = a(1 + \cos(t)) \\ y(t) = a \sin(t) \\ z(t) = 2a \sin(t/2) \end{cases} \quad (1)$$

6.1 Auswertung der Phasenabschätzung

Für diesen Fall wurde als Vorlage eine lineare Abtastung der Viviani-Kurve, und als Testdaten eine kubische Abtastung($t \rightarrow t^3$) zu der ein gleichmäßig verteiltes Rauschen hinzugefügt wurde, verwendet. Der Zustandsraum ist hier dreidimensional, bestehend aus der Phase $p_k \in [0, 1]$, der Geschwindigkeit $v_k \in \mathbb{R}$ und der Skalierung $s_k \in \mathbb{R}$, wobei v_k und s_k so normalisiert sind, dass ein Wert von 1 jeweils der Geschwindigkeit bzw. Skalierung der Vorlage entspricht. Um einen Vergleich zu GF machen zu können, wurde $v \rightarrow \infty$ gewählt.

Im Vergleich erreichten beide Verfahren gute Abschätzungen mit durchschnittlichen Fehlern von 1,3 Abtastungen beim GVF, bzw. 2,3 Abtastungen beim GF.

Weiter wurde der Einfluss von σ auf die Abschätzungen untersucht. Die Ergebnisse sind in Graph Z dargestellt. Für alle getesteten Werte für σ lieferte GVF bessere Abschätzungen als GF, obwohl GF eine genauere Inferenztechnik verwendet. Dies liegt daran, dass GVF ein besseres kontinuierliches Modell verwendet, um die Daten wiederzuspiegeln, sowie an dem Zusammenhang zwischen Geschwindigkeit und Phase, welcher von GF ignoriert wird.

Ebenso stellte sich heraus, dass GVF bessere Abschätzungen in Abhängigkeit von der Stärke des Rauschens liefert, als GF.

6.2 Auswertung der Rotationsabschätzung

Für diesen Versuch wurde die Rotation der Geste mit der Zeit variiert. Die Drehwinkel ϕ , θ und ψ für die Drehung entlang der x , y und z -Achse wurden pro Zeitschritt berechnet durch:

$$\begin{aligned}\phi(t) &= t^2 \\ \theta(t) &= t \\ \psi(t) &= -t^{1/3}\end{aligned}\tag{2}$$

Der Zustandsraum ist hier entsprechend 5-dimensional, bestehend aus p_k , v_k , ϕ_k , θ_k und ψ_k . Es wurden die gleichen Tests durchgeführt wie bei der Phasenabschätzung. Es stellt sich heraus, dass σ wenig Einfluss auf die Genauigkeit der Abschätzung hat, aber dass bei niedrigen Werten für σ die Fehlerrate stärker schwankt als bei großen Werten.

6.3 Ergebnisse der Auswertung synthetischer Daten

In beiden Experimenten wurde deutlich, dass σ wenig Einfluss auf die Abschätzung hat. Daher ist der Algorithmus auch anwendbar, wenn es nur sehr wenig Trainingsdaten gibt. Ebenso stellte sich heraus, dass die Phasenabschätzung bei festem Rauschen und σ mit einem Durchschnittsfehler von 2,3 Abtastungen sehr genau ist. Zuletzt hat die Stärke des Rauschens einen zu erwartenden Einfluss auf die Abschätzung, wobei sie noch gut genug bleibt, dass der Algorithmus auch bei signifikantem Rauschen verwendet werden kann.

7 Nutzerstudie

7.1 Durchgeführte Experimente

Es wurde eine Nutzerstudie durchgeführt, um die Einbindung von GVF in reale Anwendungen zu untersuchen. Hierfür wurde eine Anwendung entwickelt, wo eine Gestenausführung die Wiedergabe eines vordefinierten Tons hervorruft, welcher durch Variationen in der Geste verändert wurde. Hierbei wird die Früherkennung des GVF verwendet, um den Ton abzuspielen, sobald die Wahrscheinlichkeit der Erkennung über 50% beträgt. Ab dann werden für den Rest der Geste die Abschätzung der Varianzen für die Manipulation des Tons verwendet. Hierbei sorgt eine schnellere/langsamere Ausführung der Geste für eine schnellere/langsamere Wiedergabe des Tons, größere/kleinere Gesten sorgen für eine lautere/leisere Wiedergabe und eine Drehung der Geste regelt die Cut-Off-Frequenz eines Hochpassfilters. Den Teilnehmern werden 7 Aufgaben gestellt, die alle daraus bestehen, einen bestimmten Ton abzuspielen, und ihn dann zu manipulieren. Hierbei ist zu beachten, dass Aufgabe 1 nur das Abspielen eines Tons, Aufgabe 2-4 eine globale Änderung eines Aspekts, Aufgabe 5 eine kontinuierliche Änderung eines Aspekts und Aufgabe 6 und 7 die globale Änderung zweier Aspekte beinhalten. Die Gesten werden von einem infrarotbasiertem System aufgenommen, welches Fingerbewegungen im freien Raum misst. Die Studie

wurde mit 10 Teilnehmern durchgeführt, die jede der 3 Gesten 3 Mal pro Aufgabe durchführten. Somit ergeben sich 630 Datensätze.

Literatur

1. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: A \$1 recognizer for user interface prototypes. *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, 2007.