Automatische Erkennung von Kung Fu Gesten mittels Magnetfeld- und Beschleunigungssensoren

GREGOR ENDLER

Einleitung

Automatisierte Erkennung von Gesten ist ein Gebiet mit vielen Anwendungen, beispielsweise in den Bereichen Sport, Gesundheitswesen oder Entertainment. Oft kommen dabei Infrarotkameras zum Einsatz. Diese ermöglichen eine sehr präzise Abbildung von Bewegung im Raum, sind aber in Aufbau der Kameras und Anbringen der nötigen Sensoren aufwändig und teuer. Im Gegensatz dazu sind Inertialsensoren, etwa in mobilen Geräten, günstig und weitverbreitet, aber häufig auch unpräzise und finden daher oft in einfachen Gaming- oder Freizeitanwendungen Verwendung. Es existieren allerdings auch fortgeschrittene Anwendungen, die trotz relativ einfacher Sensorik komplexe Erkennungen durchführen können.
In dieser Arbeit wird untersucht, inwiefern sich die Präzision von Infrarotsensoren mit den relativ kostengünstigen Alternativen Beschleunigungs-, Rotations- und Magnetfeldsensoren erreichen lässt. Der Anwendungsbereich ist die Erkennung von Kampfsportgesten. Dies würde beispielsweise ein System ermöglichen, das das Lernen einer Kata/Form unterstützt, indem es die Gesten der User erkennt und auf Fehler im Ablauf der Form hinweist.

Gestenerkennung: Tai Chi vs. Kung Fu

Der Begriff „Geste“ wird hier für eine Bewegung mit festgelegtem Beginn und Ende verwendet, beispielsweise ein Faustschlag im Kung Fu oder „Mähne des Wildpferds teilen“ im Tai Chi. Vorgängerstudien konnten gute Ergebnisse von bis zu 100% Erkennungsrate für Tai Chi Gesten erzielen (Kunze et al., 2006; Pirkl et al., 2008). Die dort betrachteten Gesten werden relativ langsam ausgeführt und sind daher mehrere Sekunden lang. Im Gegensatz dazu dauert eine Kung Fu Geste teils unter einer halben Sekunde. Das bedeutet, dass die Tai Chi Gesten bei gleicher Sensorik bis zu eine Größenordnung mehr Datenpunkte liefern als die kurzen Kung Fu Gesten. Im Folgenden wird untersucht, ob eine zuverlässige Erkennung von Kung Fu Gesten trotz dieses Unterschiedes möglich ist.

Material und Methoden

Sensorik

Um Beschleunigung und Rotation zu messen, kamen das XBus Master System (XMB) von XSens (www.xsens.com) und zwei bis vier daran per Kabel angeschlossene

MT9 Sensoren zum Einsatz. Die Sensoren wurden an Handgelenk und Oberarm platziert (s. Abbildung 1).

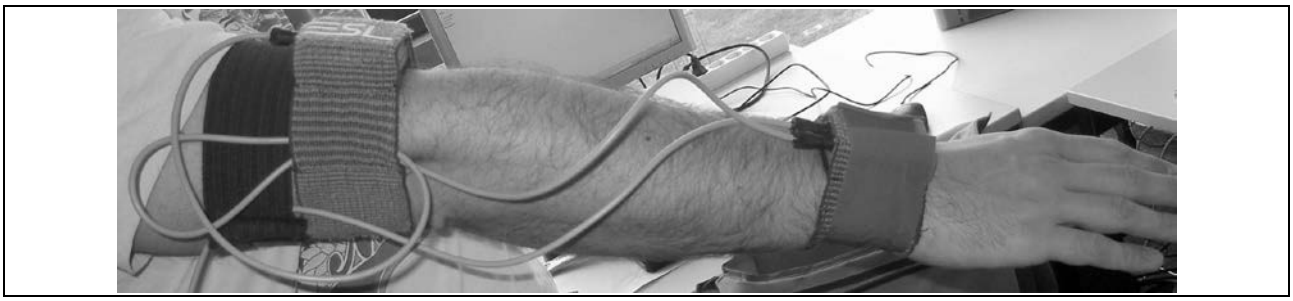

Abb. 1: Platzierung MT9 Sensoren

Jeder dieser Sensoren beinhaltet ein 3-achsiges Accelerometer, ein 3-achsiges Gyroskop und ein 2-achsiges Magnetometer, jeweils mit einer Abtastrate von 100 Hz. Die Magnetometer der MT9 Sensoren kamen nicht zum Einsatz, da sie Orientierung im Erdmagnetfeld messen, was unter Umständen fehleranfällig ist. Stattdessen wurde für Magnetfeldmessung eine Eigenentwicklung des Embedded Systems Lab Passau[1] verwendet (Pirkl et al., 2008). Hier wurden ein Magnetfeldsender und ein Magnetfeldempfänger an Brust respektive Unterarm der Testperson angebracht. Die Sensoren messen mit einer Abtastrate von 200 Hz die Magnetfeldstärke am Empfänger, und damit dessen Abstand zum Sender.
Die Daten beider Sensoren wurden per Bluetooth an einen handelsüblichen PC übertragen und im CSV Format gespeichert. Beschleunigungsdaten liegen dabei in der Einheit m/s^2 vor, Magnetfelddaten als einheitenlose relative Magnetfeldstärke (s. Abbildung 2).

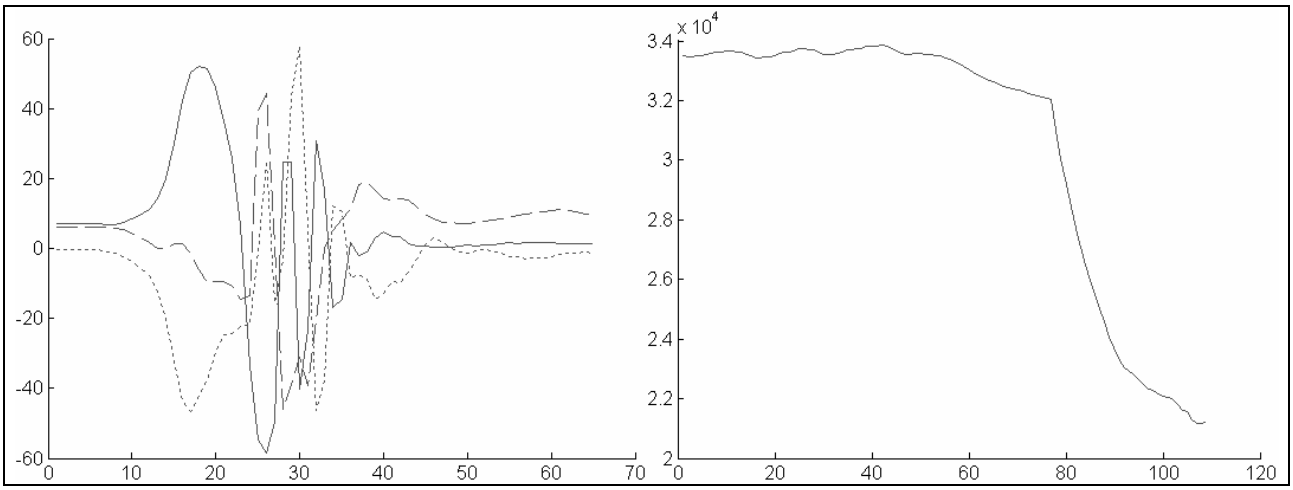

Abb. 2: gerader Faustschlag; x-Achsen: Zeit;
y-Achse links: Beschleunigung, y-Achse rechts: Magnetfeldstärke

1 Mittlerweile aufgelöst, Homepage archiviert unter
http://web.archive.org/web/20120328015408/www.esl.fim.uni- passau.de/page/home.html

Aufgenommene Gesten und Testpersonen

Sieben verschiedene Gesten wurden aufgenommen: Hoch-, Tief-, und Außenblock, gerader Faustschlag, Aufwärtshaken, Unterarmschlag und Fingerspeer. Diese Gesten wurden ausgewählt, da sie in der ersten Form des betrachteten Kung Fu Stils[2] mehrfach vorkommen, und so kein initiales Training der Testpersonen nötig war, da alle von ihnen mindestens den ersten Grad erreicht hatten und die Form und damit die Einzelgesten beherrschten. Jede der Gesten wurde mit insgesamt 6 Testpersonen je 20 mal aufgenommen. Die Kampfsport-Erfahrung der Testpersonen variierte von fortgeschrittenen Anfängern (erster Gürtel) bis hin zum Meistergrad (Schwarzgurt).

Software

Während der Aufnahme wurde jede Geste von Hand mit einem Label versehen, das Startpunkt, Ende, und Art einer Geste angibt (s. Abbildung 3). Speichern und Annotation der Daten erfolgten über die Context Recognition Toolbox (Bannach et al., 2008). Die Gestenerkennung wurde mittels zweier Techniken aus dem Machine Learning durchgeführt: Hidden Markov Models (im Folgenden HMMs) und Entscheidungsbäumen (Übersichtsartikel hierzu finden sich in (Rabiner & Juang, 1986) respektive (Witten & Frank, 2005)). Für Entscheidungsbäume wurden die Daten fensterweise verarbeitet, d.h. innerhalb fester Abschnitte der Daten wurden 30 verschiedene Kennzahlen berechnet, um diesen Abschnitt zu charakterisieren (beispielsweise Varianz, Mittelwerte und Frequenzeigenschaften der Daten im Fenster). Kennzahlberechnung und weitere nötige Vorverarbeitung der Daten erfolgte mit MathWorks MATLAB (www.mathworks.com). Die so vorverarbeiteten Daten wurden dann mit der Machine Learning Software Suite Weka (www.cs.waikato.ac.nz/ml/weka) für Entscheidungsbaumklassifikatoren verwendet. Für die Erkennung mit HMMs wurden eigene MATLAB-Skripte verwendet.

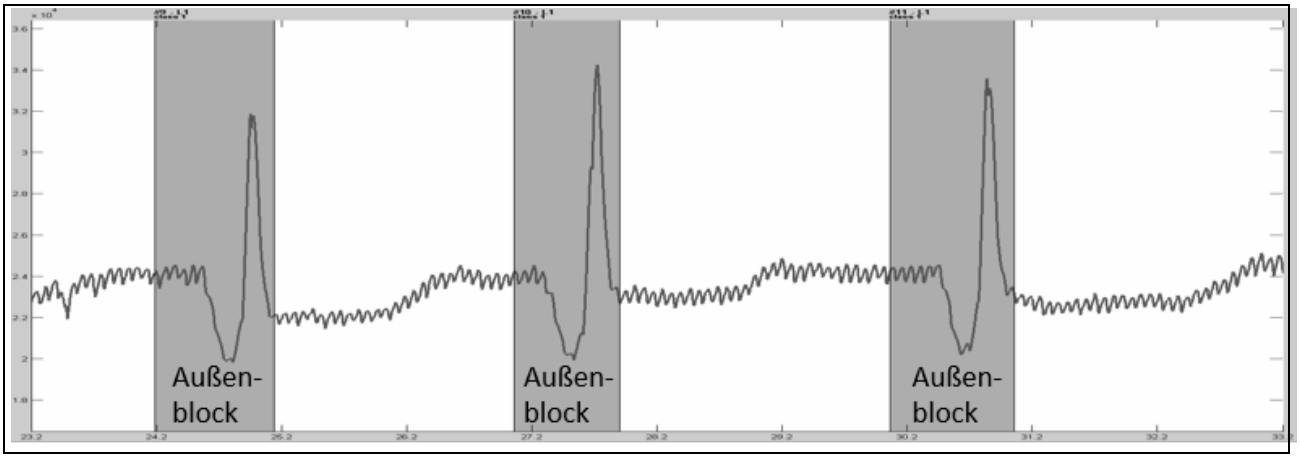

Abb. 3: Magnetfelddaten Außenblock mit Fenstern und Labels

2 Nam Wah Pai Kung Fu, *www.namwahpai-kungfu.de*

Klassifikation allgemein

Abbildung 4 zeigt den prinzipiellen Ablauf von Aufnahme und Klassifikation. Nach der Aufnahme werden die Daten in Trainings- und Testdaten eingeteilt. Aus den Trainingsdaten erstellt ein Lernalgorithmus (z.B. C4.5 für Entscheidungsbäume (Quinlan, 1993), Baum-Welch und Viterbi für HMMs (Rabiner & Juang, 1986)) ein Modell, welches die Klassifikation der Daten ermöglicht. Das Modell wird dazu auf die Testdaten angewandt, und weist jeder Instanz der Daten eine Klasse zu. Diese vorhergesagten Klassen werden dann mit den Labels der Instanzen verglichen, um so die Anzahl der korrekten Klassifikationen zu erhalten. Als Güte einer Klassifikation wird die prozentuale Anzahl korrekt klassifizierter Instanzen verwendet:

$$\frac{Anzahl\ korrekt\ klassifizierter\ Instanzen}{Gesamtzahl\ Instanzen}$$

Dieses Verhältnis wird im Folgenden „Erkennungsrate“[3] genannt.

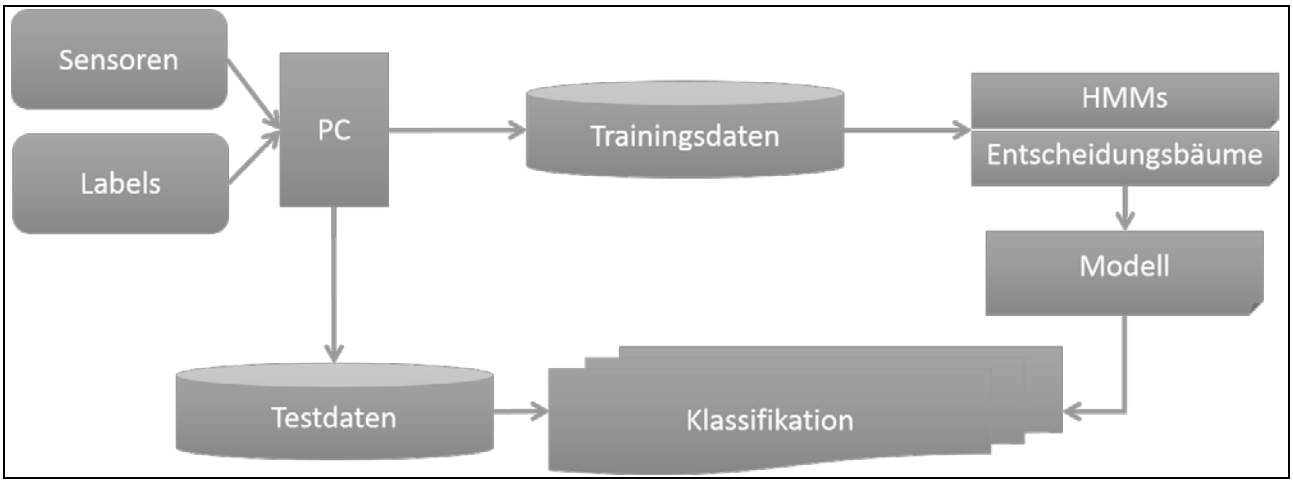

Abb. 4: Ablauf Klassifikation

Ein konstruiertes und vereinfachtes Beispiel im Fall der Kung Fu Gestenerkennung: Angenommen, in den Trainingsdaten zeichneten sich alle geraden Faustschlage durch eine maximale Beschleunigung von mindestens 40 m/s^2 aus. Das daraus gelernte Modell sei nun durch folgende Aussage beschrieben:

$$max.\ Beschleunigung > 40\ m/s^2 \Leftrightarrow Geste\ ist\ ein\ Faustschlag$$

Sind nun in den Testdaten 10 Faustschlage enthalten, von denen neun eine Beschleunigung $> 40\ m/s^2$, einer jedoch nur eine maximale Beschleunigung von 38

3 Engl. Standardbegriff: ‘accuracy’

m/s^2 aufweist, hat das Modell eine Erkennungsrate von 90%, kann also 9 von 10 Instanzen eines Faustschlags in den Testdaten korrekt erkennen.

Ergebnisse

Testpersonen-abhängige Erkennung

Für Testpersonen-abhängige Erkennung wurden die Daten jeder einzelnen Testpersonen zufällig in 2/3 Trainings- und 1/3 Testdaten aufgeteilt, die Klassifikation durchgeführt, und die Erkennungsrate protokolliert. Dies wurde für jede Person 10 mal wiederholt, um die Auswirkung der zufälligen Auswahl der Trainingsdaten zu minimieren. Durchschnittlich wurden dabei je nach Testperson Raten von 97% - 98% erzielt, die schlechtesten Raten betrugen 88% - 92%. Außerdem konnte bei jeder Testperson abhängig von der Trainingsdatenauswahl ein perfekter Klassifikator mit 100% korrekten Vorhersagen erstellt werden. Die Ergebnisse für alle 6 Testpersonen finden sich in Abbildung 5.

Testpersonen-Unabhängigkeit

Für vorangegangene Experimente wurden jeweils als Trainings- und Testdaten die Aufnahmen derselben Person verwendet. Die gelieferten Aussagen und Erkennungsraten

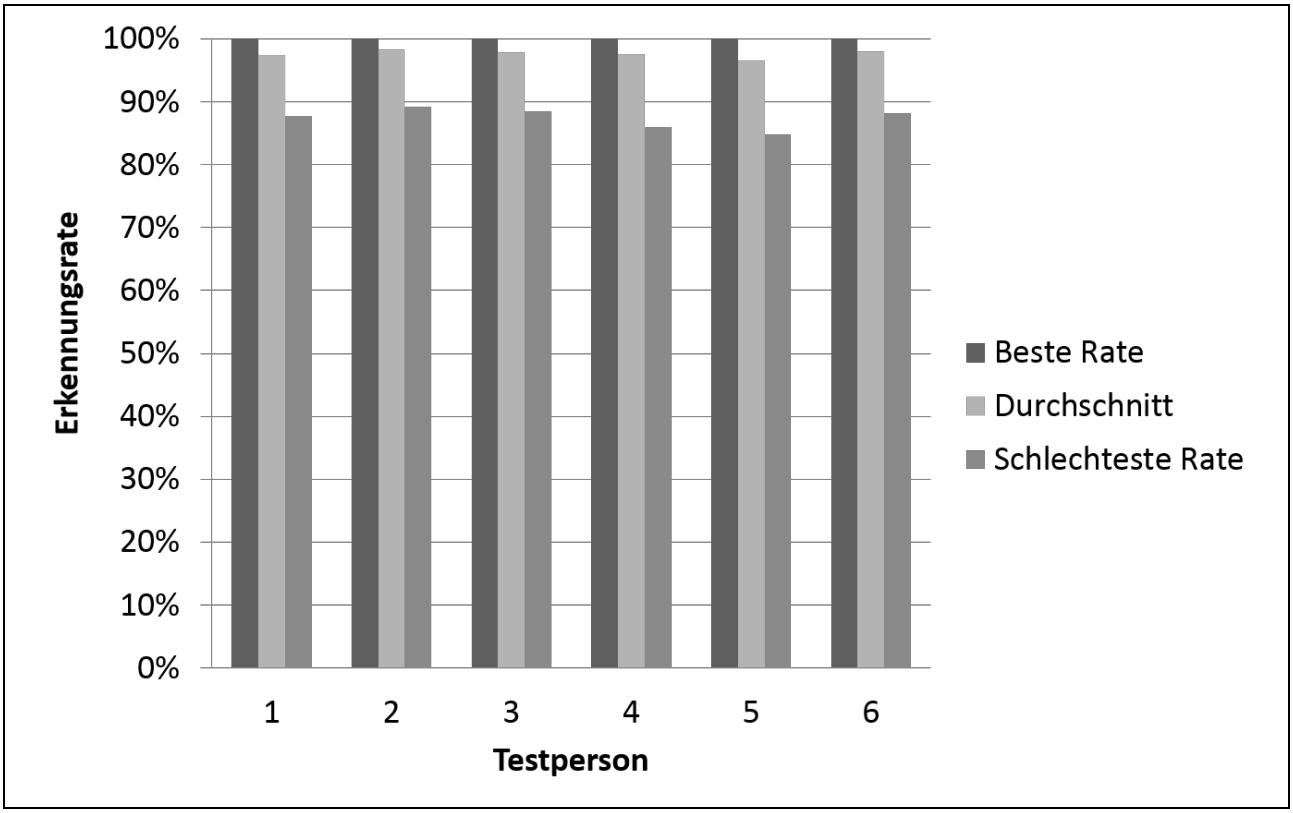

Abb. 4: Testpersonen-abhängige Erkennungsraten

gelten daher nur, falls Nutzer des Erkennungssystems selbiges vorerst auf ihre persönlichen Eigenheiten trainiert haben. Soll das System ohne Kalibrierung funktionieren, muss überprüft werden, inwiefern die Eigenheiten in der Gestenausführung der Testpersonen durch Trainingsdaten anderer Personen abgedeckt werden können. Dies testen wir durch zwei neue Zusammenstellungen der Trainings- und Testdaten:

- *Fall 1*: Test über den gesamten Datenbestand aller Testpersonen mit zufälliger Auswahl von 2/3 der Daten als Trainingsdaten.
- *Fall 2*: Auswahl der Daten von 4 Testpersonen als Trainingsdaten, Daten der verbleibenden 2 als Testdaten.

Dank der großen Anzahl von Einzelgesten und der Wiederholung des Experiments mit neuer zufälliger Verteilung kommen in Fall 1 mit größter Wahrscheinlichkeit alle Testpersonen sowohl in den Trainings- als auch in den Testdaten vor. Fall 2 ist die striktere Art der Testpersonen-Unabhängigkeit - das System hat keine Möglichkeit, die spezifischen Eigenheiten der Testdaten im Voraus zu trainieren. Die hierbei erreichten Erkennungsraten liefern eine Aussage über die Möglichkeit, ein vollständig kalibrierungsfreies Erkennungssystem zu schaffen.
Für Fall 1 konnten, ähnlich wie im Testpersonen-abhängigen Fall, sehr gute Erkennungsraten von bis zu 99% erreicht werden. Für die strikte Testpersonen-Unabhängigkeit in Fall 2 wurden um einiges niedrigere Raten festgestellt, von 48% im schlechtesten bis zu 65% im besten Fall, mit einer durchschnittlichen Erkennungsrate von rund 56% (s. Abbildung 6).

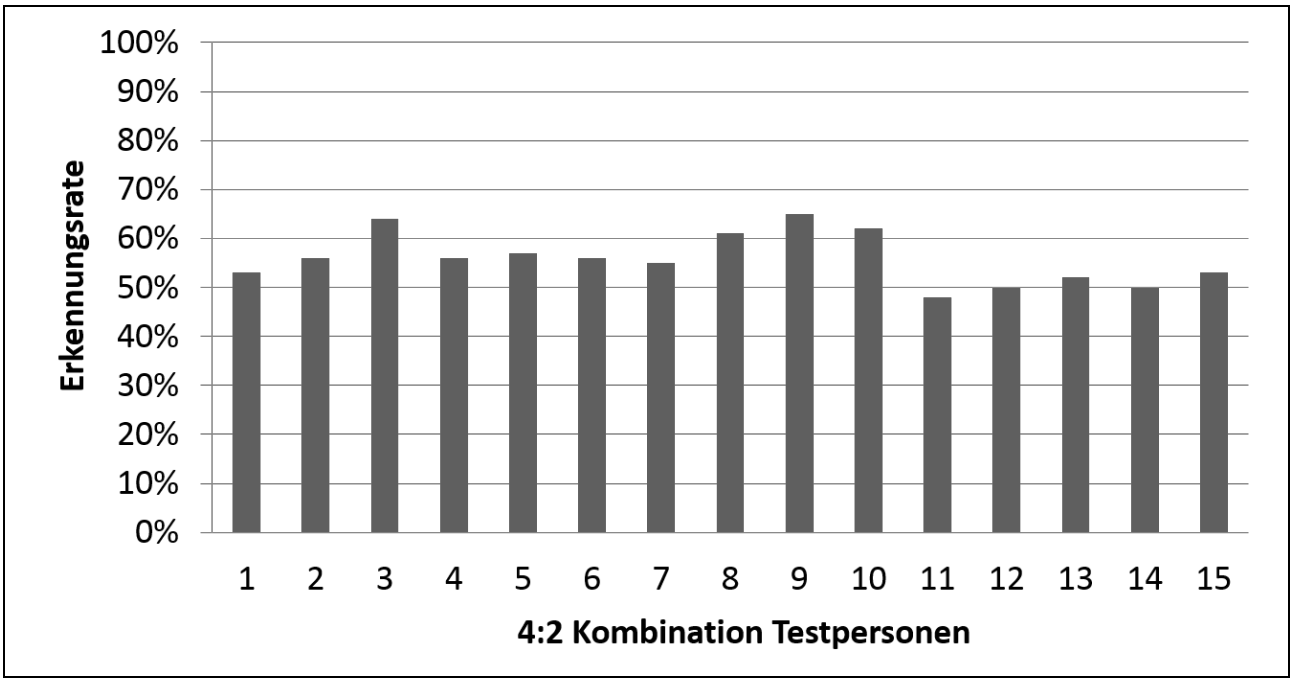

Abb. 6: Testpersonen-unabhängige Erkennungsraten für alle 4:2 Kombinationen von Trainingsdaten

Zusammenfassung

Tabelle 1 zeigt eine Zusammenfassung aller vorgestellten Ergebnisse. Technische Details, wie die vollständige Liste verwendeter Kennzahlen, einen Vergleich der verwendeten Sensoren, und Genaueres zu den verwendeten Techniken, finden sich in (Endler, 2009) (www6.cs.fau.de/people/greg/Endler2009.pdf).

Tab. 1. Zusammenfassung der Ergebnisse

Testpersonen-abhängig			**Testpersonen-unabhängig**	
Beste Rate	Durchschnitt	Schlechteste Rate	Beste Rate Fall 1	Beste Rate Fall 2
100%	97% - 98%	88% - 92%	99%	65%

Diskussion

Die Ergebnisse zeigen, dass mit den verwendeten Techniken und Daten zuverlässige Erkennung im Testpersonen-abhängigen Fall möglich ist. Die Stabilität der Ergebnisse im Vergleich der Testpersonen lasst dabei den Schluss zu, dass dies auch allgemein der Fall ist.
Für unkalibrierte Systeme, also für den Testpersonen-unabhängigen Fall, sind die Erkennungsraten jedoch noch nicht ausreichend. Fall 1 liefert, ähnlich wie die Personen- abhängige Erkennung, nahezu perfekte Ergebnisse. Das System verliert also durch heterogene Trainingsdaten nicht die Fähigkeit, die gelernten Eigenheiten zu unterscheiden. In Fall 2 sinken die Erkennungsraten jedoch stark ab. Überraschend ist dabei der hohe Unterschied zwischen beiden Fällen - von fast perfekter Erkennung in Fall 1 zu bestenfalls 65% Erkennungsrate in Fall 2. Auch wenn dies immer noch eine signifikante Verbesserung zu einem zufälligen (ratenden) Klassifikator[4] darstellt, ist diese Rate für praktische Anwendung quasi unbrauchbar. Der wahrscheinlichste Grund für dieses Ergebnis ist „overfitting“: Der Klassifikator lernt die Eigenheiten der 4 Testpersonen in den Testdaten zu genau, was seine Fähigkeit zu generalisieren einschränkt. Eine größere Anzahl von Testpersonen für die Trainingsdaten könnte dies beheben. Um diese Theorie zu überprüfen sind weitere Aufnahmen mit neuen Testpersonen nötig.
Eine weitere interessante Fragestellung ist, ob sich anhand der Daten Rückschlüsse auf die „korrekte“ Ausführung einer Geste oder auf die Kampfsport-Erfahrung der Testperson ziehen lassen. Ein Indiz, das Letzteres unter Umständen möglich ist, ist die Tatsache, dass die einzige Testperson mit Meistergrad auch die einzige war, deren Gesten regelmäßig die maximal messbare Beschleunigung von 60 m/s^2 überstiegen. Auch dies bedarf jedoch zukünftiger Überprüfung.

4 Ein zufälliger Klassifikator würde bei 7 Gesten rund 14,3% Erkennungsrate erreichen.

Literatur

Bannach, D., Lukowicz, P., & Amft, O. (2008). Rapid prototyping of activity recognition applications. Pervasive Computing, IEEE, 7(2):22–31.

Endler, G. (2009). Activity recognition for martial arts applications. Diplomarbeit, Universitat Passau.

Kunze, K., Barry, M., Heinz, E. A., Lukowicz, P., Lukowicz, P., Majoe, D., & Gutknecht, J. (2006). Towards recognizing tai chi - an initial experiment using wearable sensors. In Applied Wearable Computing (IFAWC), 2006, pages 1–6.

Pirkl, G., Stockinger, K., Kunze, K., & Lukowicz, P. (2008). Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton. In Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on, pages 47–54.

Quinlan, J. R. (1993). C4. 5: Programs for Machine Learning, volume 1. Morgan Kaufmann.

Rabiner, L. & Juang, B. (1986). An introduction to hidden markov models. ASSP Magazine, IEEE, 3(1):4–16.

Witten, I. H. & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.