Diss. ETH No. 16719

Low-Power Sound-Based User Activity Recognition

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by

MATHIAS STÄGER

Dipl. El.-Ing. ETH
born 11th August 1975
citizen of Zizers GR

accepted on the recommendation of

Prof. Dr. Gerhard Tröster, examiner
Prof. Dr. Paul Lukowicz, co-examiner

2006

Abstract

In the near future, tiny computers and sensors integrated into our clothes will monitor our physical condition and activities to assist us in everyday tasks: be it suggesting simple improvements in a recipe while we cook or even preventing a maintenance worker from a hazardous mistake.

Typically, activities are recognized with motion sensors. In this work, we use sound as a novel modality for activity recognition. With a microphone mounted on the wrist, e.g. inside a watch, we are able to pick up sounds that are caused by the user or occur in close proximity to the user's hand. Thus, we can detect hand movements which generate a sound, like switching on an appliance or performing a noisy manual task.

Activity recognition systems are usually tuned to deliver high recognition rates. However, size and battery capacity of wearable systems impose limitations to the type and amount of sound processing that can be done. The challenge is to carefully select the parameters of the recognition process to achieve on the one hand high recognition rates and on the other hand low power consumption. Thus, our work describes advances towards the development of a power optimized recognition system: a system that is tuned during the training phase of the recognition process to represent a trade-off between power consumption and recognition accuracy.

We discuss advantages and limitations of the sound-based activity recognition approach. We show that reasonable scenarios exist where sound is a useful means to detect activities. We present case studies in which we recorded data with wrist worn microphones for four scenarios: operating kitchen appliances, performing manual tasks in a wood workshop, operating office appliances and being outdoors/using public transport. These recordings are used to benchmark various sound processing algorithms.

To make a low-power, sound-based activity recognition system feasible, we work with a frame-based method – in contrast to a continuous recognition. This requires segmentation procedures which partition the data stream into potentially interesting segments. One technique proposed here is based on the difference of the signal amplitudes of a wrist and a chest worn microphone. Furthermore, we analyze the recognition process and discuss how recognition accuracy is affected by various parameters like sampling frequency, sampling duration, number of frames, choice of features and classifiers. In addition to sound, we investigate another sensor modality: acceleration. Different methods to fuse data from two sensors are explored.

To estimate power consumption, we use a hardware platform containing a microcontroller, a microphone and accelerometers. Based on power measurements of selected operating points and a simplified model of the hardware platform, we synthesize its power consumption for various sampling frequencies, sampling durations and signal processing algorithms running on the microcontroller.

Finally, we combine the two conflicting metrics 'recognition rate' and 'power consumption' to a pareto plot. With the data from our case studies we demonstrate that the methods proposed in this work lead to improvements in battery lifetime by a factor of 2 to 4 with only little degradation in recognition performance. To conclude, a wrist worn sensor node recognizing 5 different kitchen appliances with 94% accuracy consumes as little as 0.55 mW. This allows to power it for 42 days with a 2 cm^3 lithium-ion battery.

Zusammenfassung

In naher Zukunft werden winzige, in Kleidung integrierte, Computer und Sensoren unseren Gesundheitszustand und unsere Aktivitäten überwachen, um uns bei alltäglichen Tätigkeiten zu unterstützen: sei es um Verbesserungsvorschläge an einem Kochrezept, das wir gerade ausprobieren, anzubringen oder um einen Servicetechniker vor einem gefährlichen Fehlgriff zu warnen.

Üblicherweise werden zur Erkennung von Aktivitäten Bewegungssensoren verwendet. In dieser Arbeit wählen wir einen neuartigen Ansatz und verwenden Geräusche, um Aktivitäten zu erkennen. Mit einem Mikrophon, das am Handgelenk befestigt ist, z.B. in einer Uhr, können wir Geräusche erfassen, welche durch den Benutzer verursacht werden oder in der näheren Umgebung seiner Hand auftreten. Dadurch können wir diejenigen Handbewegungen detektieren, die ein Geräusch auslösen, wie zum Beispiel das Einschalten eines Geräts oder eine laute manuelle Tätigkeit.

Systeme zur Erkennung von Aktivitäten sind üblicherweise auf eine hohe Erkennungsrate ausgerichtet. Allerdings beschränken Grösse und Batteriekapazität von tragbaren Computern die Art und Weise, wie Geräuscherkennung durchgeführt werden kann. Deshalb ist es eine Herausforderung, die Parameter des Erkennungsprozesses so zu wählen, dass einerseits eine hohe Erkennungsrate und andererseits ein geringer Leistungsverbrauch erzielt wird. In dieser Arbeit streben wir ein Aktivitäten-Erkennungs-System an, das einen Kompromiss zwischen Leistungsverbrauch und Erkennungsrate darstellt.

Wir diskutieren die Vorteile und Einschränkungen eines Aktivitäten-Erkennungs-Systems, das sich auf Geräusche stützt. Wir präsentieren Fallstudien, in welchen wir mit am Handgelenk getragenen Mikrofonen für vier Szenarien Daten aufgenommen haben: Ein- und Ausschalten von Küchengeräten, verschiedene manuelle Tätigkeiten im Bereich von Holzverarbeitung, Tätigkeiten im Büro und Geräusche die im Freien respektive in öffentlichen Verkehrsmitteln vorkommen. Die Aufnahmen werden verwendet, um Leistungsvergleiche zwischen verschiedenen Algorithmen, die Geräusche verarbeiten, durchzuführen.

Um ein Aktivitäten-Erkennungs-System mit einer niedrigen Leistungsaufnahme zu realisieren, arbeiten wir mit einer Methode, die einzelne 'Frames' oder Fenster analysiert – im Gegensatz zu einer kontinuierlichen Erkennung. Dazu werden Verfahren notwendig, die den Datenstrom in potenziell interessante Segmente aufteilen können. Eine Technik, die hier

vorgestellt wird, basiert auf dem Lautstärkeunterschied zwischen einem Mikrofon am Handgelenk und an der Brust. Des Weiteren untersuchen wir den Erkennungsprozess und diskutieren dabei, inwiefern die Erkennungsrate von Parametern wie der Abtastfrequenz, der Abtastdauer und der Wahl der Erkennungsmerkmale und Klassifikatoren abhängt. Zusätzlich zum Mikrofon betrachten wir die Verwendung eines zweiten Sensors: eines Beschleunigungssensors. In diesem Zusammenhang beschäftigen wir uns mit unterschiedlichen Verfahren, um die Daten von den zwei Sensoren zusammen zu führen.

Um den Leistungsverbrauch abzuschätzen, verwenden wir eine Hardware, die einen Mikrokontroller, ein Mikrofon und Beschleunigungssensoren enthält. Basierend auf Stromverbrauchsmessungen von ausgewählten Operationsmodi und einem vereinfachten Modell der Hardware berechnen wir deren Leistungsaufnahme für verschiedene Abtastfrequenzen, Abtastlängen und signalverarbeitende Algorithmen.

Schlussendlich kombinieren wir die zwei Metriken 'Erkennungsrate' und 'Leistungsverbrauch' zu einer Pareto Grafik. Mit Hilfe unserer Fallstudien zeigen wir, dass die in dieser Arbeit vorgestellten Methoden die Batterielebensdauer um Faktor 2 bis 4 verlängern können bei beinahe gleichbleibender Erkennungsrate. Letztendlich verbraucht ein am Handgelenk getragenes System, das 5 verschiedene Küchengeräte anhand des Geräusches mit einer Wahrscheinlichkeit von 94% erkennt, nur 0.55 mW. Dadurch kann es für 42 Tage mit einer nur $2\,\mathrm{cm}^3$ grossen Litium-Ionen Batterie betrieben werden.