

Erkennung von Landmarken für Roboternavigation mit Luftultraschall und neuronalen Netzen – Feature Engineering und Vergleich

Patrick K. Kroh¹, Ralph Simon², Stefan J. Rupitsch¹

¹Lehrstuhl für Sensorik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Paul-Gordan-Str. 3/5, 91052 Erlangen, Deutschland

²Department of Ecological Science, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, Niederlande

Zusammenfassung

Ultraschallbasierte Sensoren kommen bei mobilen Robotern und im Automobilbereich häufig zur Abstandsmessung von umliegenden Objekten zum Einsatz. Das akustische Messprinzip unterscheidet sich physikalisch von den Prinzipien anderer geläufiger Sensoren zur Umfelderkennung von Robotern (z. B. Lidar, Radar, Kamera etc.) und folglich können entsprechende Sensorsysteme mit Ultraschallsensoren komplementär ergänzt werden. Es wird dadurch eine Reduktion der Fehleranfälligkeit und somit Erhöhung der Zuverlässigkeit von Roboternavigationssystemen möglich. In unserer Arbeit beschäftigen wir uns daher mit der Identifizierung und Klassifizierung von Ultraschall-reflektierenden Objekten, welche als ortsfeste Referenzpunkte (Landmarken) für die Navigation herangezogen werden können. Hierzu werden künstliche neuronale Netze als Klassifizierer mit unterschiedlichen Merkmalsvektoren (Features) trainiert und untersucht. Es hat sich gezeigt, dass mit den entwickelten Features, auf Basis der Target Strength der Ziele, verschiedene geometrische Formen zu einem hohen Grad und auch bei zusätzlichem Rauschen erfolgreich identifiziert werden können.

Keywords: Landmarkenerkennung, Sonarmessung, Zielerkennung, Roboternavigation, Luftultraschall

Einleitung

In Anwendungsgebieten wie der mobilen Robotik und im Automobilbereich, wird Ultraschallsensorik in Luft häufig eingesetzt, um das Vorhandensein von nahegelegenen Hindernissen zu detektieren und deren Abstand zu bestimmen [1-4]. Im Gegensatz zu anderen Sensorprinzipien zur Umfelderkennung erfolgt diese dabei unabhängig von den optischen und elektromagnetischen Eigenschaften der zu detektierenden Hindernisse. Folglich ist auch eine Erkennung von optisch reflektierenden und transparenten Zielen mit geringem Radarquerschnitt möglich (z. B. Objekte aus transparenten Kunststoffen oder Hindernisse im Dunkeln), welche häufig eine Herausforderung für Systeme wie Kameras, RADAR (Radio Detection And Ranging) und LIDAR (Light Detection And Ranging) darstellen. Im Vergleich zu den genannten Prinzipien weisen ultraschallbasierte Systeme allerdings normalerweise eine geringe Winkelauflösung bei reinem Puls-Echo-Betrieb auf, wodurch eine Identifizierung der Umgebung sowie einzelner reflektierender Objekte anhand deren Geometrie nur bei ver-

hältnismäßig großen Abmessungen möglich ist. Im Rahmen unserer Arbeit entwickeln wir daher ein System zur Identifikation und Ortung von spezifischen schallreflektierenden Objekten (Ziele) direkt anhand ihrer Echosignale. Die Objekte können dabei auch deutlich kleiner sein (charakteristische Abmessungen von ca. 100 mm). Dies kann besonders nützlich sein, wenn ein Roboter in einer Umgebung navigieren muss, die sich ändern kann. So könnten beispielsweise gut zu detektierende Objekte an festen Punkten angebracht werden und als ortsfeste Referenzpunkte dienen (Landmarken) [5,6]. Dies kann z. B. der Fall sein, wenn Roboter in Umgebungen agieren müssen, in welchen sich auch Menschen aufhalten, wie in den Bereichen der Assistenzrobotik und der Mensch-Roboter-Kooperation in der industriellen Fertigung [7,8]. Hierbei ist vor allem noch durch Integration weiterer komplementärer Sensorprinzipien mit Sensordatenfusion eine deutliche Erhöhung der Zuverlässigkeit der Navigation zu erwarten.

In dem vorliegenden Manuskript wird eingegangen auf die grundlegende Methodik zur Identifikation und Klassifizierung von Landmar-

ken. Der Fokus liegt dabei auf den Klassifizierern sowie der Entwicklung von Merkmalsvektoren, mit denen die Klassifizierer effizient arbeiten können. Es werden unter anderem verschiedene akustische Anregungssignale und der Einfluss von Rauschen untersucht. Als Klassifizierer kommen künstliche neuronale Netze (KNNs) zum Einsatz, da sich diese aufgrund ihrer Eigenschaften in bereits ähnlich gelagerten Aufgabenstellungen als vorteilhaft erwiesen haben. Dies wurde beispielsweise in Arbeiten von Eliakim et al., Ayulu et al. und Dmitrieva et al. gezeigt, welche allerdings andere Features entwickelt sowie verwendet haben, unterschiedliche Signale eingesetzt haben und auch nicht den Einfluss von Rauschen untersucht haben [9-12].

Messaufbau und -ablauf

Echosignale wurden in einem schalltoten Raum für verschiedene Ziele und Positionen aufgenommen sowie anschließend an einem handelsüblichen Desktop-Computer weiterverarbeitet und klassifiziert.

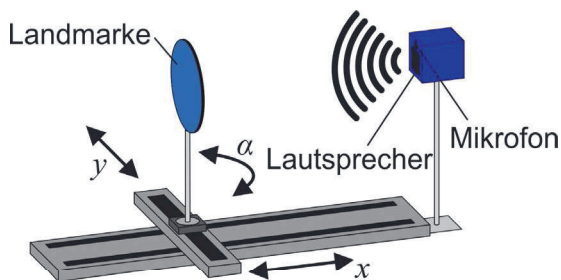


Abb. 1: Versuchsaufbau mit Ultraschalllautsprecher, Landmarke, Mikrofon, translatorischen Verfahrenseinheiten (x , y) und Rotationseinheit (α).

Hierzu wurden die Ziele mit 1 m Abstand vom Boden auf einer Stange montiert, welche motorisiert in zwei translatorischen Richtungen (x , y) verfahren und auch rotiert werden kann. Die Ultraschallsignale wurden von einem breitbandigen kapazitiven Ultraschalllautsprecher (Senscomp 7000, 20 kHz bis > 150 kHz, > 85 dB re $20\mu\text{Pa}$) emittiert, von den Zielen reflektiert und die Echosignale mit einem $\frac{1}{4}$ "-Messmikrofon (Brüel & Kjær 4939-A-011) mit Vorverstärker (G.R.A.S. 12AK) aufgenommen. Als Signale kamen breitbandige (150 bis 20 kHz) und schmalbandige (52 bis 48 kHz) rechteckförmige Downchirp-Signale mit einer Dauer von 1 ms zum Einsatz [13]. Der Ultraschalllautsprecher und das Mikrofon sind möglichst nahe beieinander montiert (20 mm Mittenabstand) und so angebracht, dass sich die Mitte des Lautsprechers auf Höhe der Mitte der Ziele befindet. Eine Messdatenerfassungskarte (National Instruments NI-USB-6356,

1,25 MS/s, 16 bit) wurde eingesetzt für das Generieren der elektrischen Anregungssignale und das Analog-Digital-Umsetzen der aufgenommenen Echosignale. Der verwendete Lautsprecher erfordert verhältnismäßig hohe Anregungsspannungen im Bereich von 0 bis 400 V. Es wurde daher ein spezieller Verstärker für diesen gebaut und eingesetzt. Die Ziele wurden mit einem 3D-Drucker (Ultimaker 2) aus ABS-Kunststoff hergestellt.

Für jedes Zielobjekt wurden Echosignale für alle Kombinationen aus x -Positionen 0,5 m bis 1,8 m (in 0,1 m-Schritten), y -Positionen -0,15 m bis 0,15 m (in 0,05 m-Schritten) und Winkeln α von -60° bis 60° (in 15° -Schritten) aufgenommen. Die aufgenommenen Signale wurden weiterverarbeitet, Merkmale aus diesen extrahiert und anschließend für das Training der KNNs herangezogen. Es wurden dabei immer Echosignale für drei benachbarte y -Positionen berücksichtigt. Dies entspricht z. B. den Aufnahmen, die bei Vorbeifahrt eines Roboters an einem Hindernis gemacht würden.

Target Strength und Ziele

Ein wichtiges Merkmal von Sonarzielen ist deren sogenannte Target Strength, welche ein Maß für die reflektierte akustische Intensität in Bezug auf die einfallende akustische Intensität darstellt. Der Begriff der Target Strength kann sich auf die gesamte Intensität beziehen, oder aber auch in Abhängigkeit der Signalfrequenz ausgedrückt werden. Es werden daher nachfolgend auch die Begriffe der gesamten Target Strength und der spektralen Target Strength verwendet.

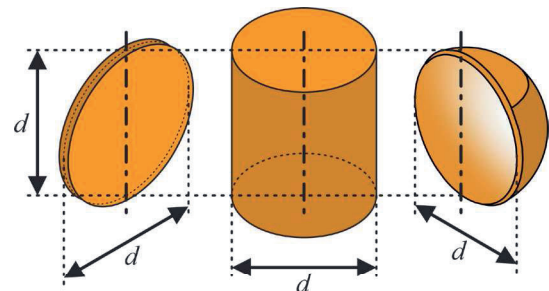


Abb. 2: Zielgeometrien mit charakteristischer Abmessung d ; Scheibe, Zylinder und hohle Halbkugel.

Untersucht man die Target Strength verschiedener Objekte, so fällt auf, dass diese auch vom Beschallungswinkel abhängen kann. Es zeigen sich dabei deutliche Unterschiede für flache, konkave sowie konvexe Objekte, falls die Objekte ausreichend groß sind im Vergleich zur Wellenlänge der eingesetzten akustischen Signale. Für die Untersuchungen der vorliegenden Arbeit wurden möglichst einfache, grundlegende Geometrien ausgewählt,

welche die zuvor genannten Eigenschaften aufweisen: kreisförmige Scheiben (flach), Zylinder (konvex) und hohle Halbkugeln (konkav). Für die Trainingsdatensätze der KNNs wurden Ziele in jeweils zwei Größen (Durchmesser d jeweils 60 mm und 100 mm) eingesetzt, um eine Generalisierung bezüglich der Form zu ermöglichen.

Bei flachen Objekten zeigt sich ein deutliches Maximum bei senkrechter Einstrahlung und ein starker Abfall mit zunehmendem Winkel. Folglich wird bereits bei leichter Verkippung ein Großteil der Energie nicht mehr zum Ausgangsort zurück reflektiert. Dies ist deutlich zu erkennen im sogenannten akustischen Fingerabdruck, welcher die spektrale Target Strength in Abhängigkeit des Verdrehwinkels eines Objekts angibt (vgl. Abbildung 3).

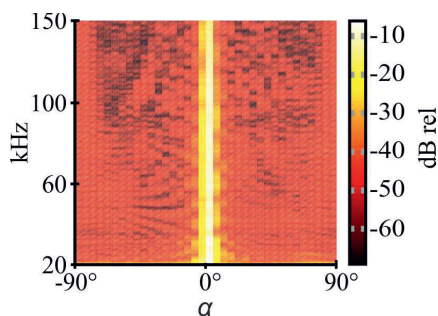


Abb. 3: Akustischer Fingerabdruck einer Scheibe (spektrale Target Strength für verschiedene Drehwinkel α).

Bei konvexen Objekten ist über einen relativ großen Winkelbereich ein Echo detektierbar, allerdings wird auch immer ein Teil der einfallenden Welle in andere Richtungen reflektiert. Es folgt daher, dass die Target Strength gleichmäßiger verteilt ist über den Verdrehwinkel, aber nicht so hoch ist wie bei senkrechtem Einfall bei einem flachen Objekt. Bei einem Zylinder ist die Target Strength optimalerweise unabhängig vom Verdrehwinkel (vgl. Abbildung 4).

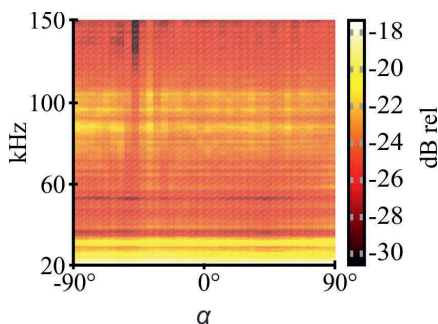


Abb. 4: Akustischer Fingerabdruck eines Zylinders.

Bei konkaven Objekten kann es zur Fokussierung kommen, wodurch eine relativ hohe Tar-

get Strength für bestimmte Winkel auftreten kann. Außerdem sind Mehrfachreflexionen innerhalb eines Objektes möglich, die sich überlagern und durch Interferenz zu einer deutlichen Frequenzabhängigkeit der Target Strength führen [14]. Die angesprochenen Zusammenhänge sind deutlich in Abbildung 5 zu sehen.

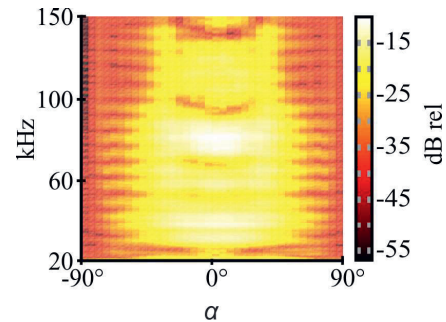


Abb. 5: Akustischer Fingerabdruck einer Halbkugel.

Die akustischen Fingerabdrücke in den Abbildungen wurden mit einem separaten Messverfahren ermittelt, wie in [13] dargestellt.

Künstliche Neuronale Netze

Künstliche neuronale Netze sind Nervensystemen bei realen Lebewesen nachempfunden und werden häufig als Klassifizierer im Bereich der Mustererkennung und der künstlichen Intelligenz eingesetzt. Populäre Beispiele sind dabei die Handschriften- und Spracherkennung. Hierbei zeigen sich besonders die vorteilhaften Eigenschaften künstlicher neuronaler Netze, da diese von gegebenen begrenzten Datensätzen (i) lernen und (ii) generalisieren können sowie (iii) mit verrauschten und verzerrten Bildern und Signalen arbeiten können [15-17]. Aus (i) folgt außerdem, dass eine explizite Modellierung des Problems selbst nicht zwingend notwendig ist. Dies ist besonders vorteilhaft für Aufgabenstellungen, bei welchen bestimmte Merkmale nur aufwändig zu analysieren und modellieren sind, wie bei den zuvor genannten Beispielen.

Die Daten, die am Eingang eines KNNs (Input Layer) anliegen, werden allgemein als Merkmalsvektor oder Featurevektor bezeichnet. Es kann sich dabei um Rohdaten (Bilder, Signalverläufe etc. bei Deep Learning) oder auch um vorverarbeitete Daten handeln, die dann auch als Merkmale oder Features bezeichnet werden. Eine solche Vorverarbeitung dient häufig dazu, eine alternative und ggf. auch kompaktere Darstellung der Daten zu finden, bei welcher die bedeutenden Informationsanteile leichter für die KNNs zu erkennen sind, sodass ein effizienteres Lernen und eine höhere Performanz erzielt werden können. Die Entwicklung

entsprechender Darstellungen unter Berücksichtigung von Wissen über die Problemdomäne wird auch als Feature Engineering bezeichnet und wurde im Rahmen der vorgestellten Arbeit für die Problemstellung der Klassifizierung von reflektierenden Objekten für Ultraschall durchgeführt.

Vorteilhaft beim Einsatz von Engineered Features ist, dass häufig deutlich kleinere Trainingsdatensätze erforderlich sind im Vergleich zu Deep Learning, was besonders bei speziellen technischen Anwendungen wünschenswert ist, da hier häufig nur begrenzte Datensätze vorliegen. Der geringere Bedarf an Trainingsdaten ist darin begründet, dass bei Deep Learning zunächst bedeutungsvolle Features von den Netzen, ausgehend von Rohdaten, selbst gelernt werden müssen (Feature Learning). Diese können bei Engineered Features durch eingebrachtes Domänenwissen bereits vorgegeben sein. Allerdings erfolgt hier das Lernen nur auf den vorab definierten Features, was dazu führen kann, dass bedeutungstragende Informationen aus Rohdaten unberücksichtigt bleiben, da diese explizit erfasst/modelliert werden müssen.

Als Netze kommen bei uns mehrschichtige Perzeptron-Netze zum Einsatz (Abbildung 6), da diese relativ stark abhängig sind von der Qualität der eingesetzten Features und somit anhand der erzielbaren Klassifizierungsperformanz auf die Qualität der entwickelten Features zurückgeschlossen werden kann. Die eingesetzten Netze besitzen vier verdeckte Schichten (Hidden Layers) mit jeweils zehn, fünf, fünf und drei Knoten. Im Vergleich zu einschichtigen Netzwerken ist durch die Tiefe eine Ableitung von abstrakten, komplexen Merkmalen durch die Netzwerke selbst leichter möglich. Außerdem ist das Risiko einer Überanpassung an die Trainingsdatensätze relativ gering, da die Anzahl der Knoten auch niedriger ist als bei anderen Netzwerken für ähnliche Aufgaben [9-12]. Unter Überanpassung ist zu verstehen, dass ein Netz vielmehr die Trai-

ningsdaten auswendig lernt anstatt davon ausgehend zu generalisieren. Dies hat dann zur Folge, dass unbekannte Daten sehr schlecht klassifiziert werden. Dazu kann es kommen, falls die Anzahl der Knoten in einem Netzwerk zu groß ist im Verhältnis zur Menge der vorhandenen Trainingsdaten. Um Überanpassung zu entdecken, wird folglich die Performanz eines Netzes mit Testdatensätzen ermittelt, die nicht Teil der Trainingsdaten sind.

Die KNNs wurden hier mit einem sogenannten Scaled Conjugate Gradient Backpropagation-Algorithmus trainiert, bei welchem noch eine Kreuzvalidierung auf einem separaten Datensatz durchgeführt wird. Die vorhandenen Messdaten wurden zu 20 % auf den Trainingsdatensatz, zu 10 % auf den Validierungsdatensatz und zu 70 % auf den Testdatensatz aufgeteilt. Der vergleichsweise geringe Anteil für den Trainingsdatensatz wurde gewählt, um eine Überanpassung aufgrund von Symmetrien im Aufbau und somit redundanten Echo-signalen zu vermeiden.

Für alle Zielgeometrien (Halbkugel, Zylinder, Scheibe) gibt es jeweils eine eigene Klasse für die Klassifikation (Output Layer) sowie eine zusätzliche Klasse für den Fall, dass kein Ziel (kZ) vorliegt. Zur Beurteilung der Performanz von KNNs können verschiedene Maße herangezogen werden. Hier wird die Korrektklassifikationsrate (Accuracy) herangezogen, welche gegen 100 % streben sollte. Die Korrektklassifikationsrate bezeichnet das Verhältnis der Anzahl aller korrekten Klassifizierungen zur Gesamtanzahl an Testbeispielen.

Weitere Maße (z. B. Precision, Recall, F1-Score [18-20]) wurden im Rahmen der durchgeführten Tätigkeiten herangezogen, allerdings führten diese zu den gleichen Ergebnissen wie die Korrektklassifikationsrate und werden daher in diesem Manuskript nicht weiter aufgeführt. In der Veröffentlichung [13] der Autoren wird allerdings auf die Maße in einem erweiterten Kontext eingegangen.

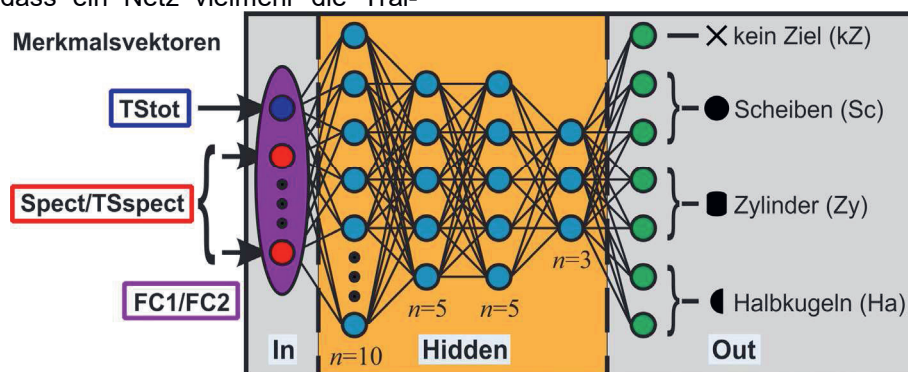


Abb. 6: Künstliches neuronales Netz mit Features (T_{Stot} , $Spect$, TS_{spect} , $FC1$, $FC2$), Eingangsschicht, verdeckten Schichten (Hidden) und Ausgangsschicht (Out) mit Zielklassen.

Features

Mit den KNNs wurden verschiedene Featurevektoren untersucht. Diese basieren unter anderem auf Spektrogrammen der Echosignale (*Spect*), Schätzungen der gesamten (*TStot*) sowie der spektralen (*TSspect*) Target Strength und Kombinationen aus den zuvor genannten Merkmalen ($FC1 = TStot$ und *Spect*, $FC2 = TStot$ und *TSspect*).

Bei der Berechnung der einzelnen Merkmale wurde wie folgt vorgegangen (vgl. auch Abbildung 7). Zunächst erfolgt die Berechnung der Kreuzkorrelationsfunktion des Echosignals mit einer vorhandenen Aufnahme des akustischen Sendesignals, die bei 1 m Mikrofonabstand aufgenommen wurde. Dadurch werden unkorrelierte Signalanteile (z. B. Rauschen und Signale anderer Quellen) herausgefiltert und zudem sind einzelne Reflexionen in der Kreuzkorrelationsfunktion deutlicher erkennbar als im Rohsignal. Im Idealfall besteht nämlich das gefilterte Signal aus einer Überlagerung von um die jeweilige Ausbreitungszeit verschobenen und skalierten Autokorrelationsfunktionen des Sendesignals, die den einzelnen Reflexionen entsprechen und in ihrer Breite abhängig sind von der Bandbreite des Sendesignals. Diese Technik ist im Bereich von Sonar- und Radarsystemen bekannt als sogenannte Pulskompression [21-24].

Aus den pulskomprimierten Signalen wird nun ein Bereich (Region of Interest, ROI) ausgeschnitten, der zentriert ist um den Spitzenwert, der dem Echo des zu klassifizierenden Ziels entspricht. Für die Klasse kZ werden zufällig Spitzenwerte ausgewählt, die sich außerhalb der ROIs von den bekannten Zielen befinden. Für *Spect* werden ROIs entsprechend in den Spektrogrammen der Rohsignale ausgewählt. Die Dauer der ROIs beträgt 2 ms. Die an-

schließenden Berechnungen erfolgen dann nur für die ROIs, sodass die KNNs nicht lernen müssen die Signalanteile auszublenden, die nicht im Zusammenhang mit den zu klassifizierenden Zielen stehen.

TStot und *TSspect* werden aus den pulskomprimierten Echosignalen ermittelt, wobei für *TStot* nur der Spitzenwert des Pulses relevant ist, der von der Reflexion an einem Ziel herührt. Die konkreten formelmäßigen Zusammenhänge und deren Herleitung sind gegeben in [13]. Die Herleitung beruht unter anderem auf der Sonargleichung [25] sowie allgemeinen Beziehungen zwischen Signalleistung, Fouriertransformation und Korrelationsfunktionen. Außerdem wird von der Annahme ausgegangen, dass der Signalpegel entlang des Ausbreitungsweges nur aufgrund von geometrischer Aufweitung der Welle abnimmt. Für *TSspect* wird außerdem anstatt einer einzelnen Fouriertransformation über die ROI in einem pulskomprimierten Signal, eine Kurzzeitfouriertransformation (STFT) mit einer Fensterlänge von 1024 Samples und einem Überlappungsbereich von 50 % durchgeführt. Dies erfolgt in Analogie zu Spracherkennungssystemen, bei welchen sich der Einsatz von STFTs als vorteilhaft gezeigt hat und mittlerweile den Regelfall darstellt [26-30]. Die gleichen Parameter wurden auch für die STFT der Roh-Echosignale für das Feature *Spect* verwendet. Für jedes Feature werden die Ergebnisse der Berechnungen von drei Positionen in einem 1D-Merkmalvektor zusammengefasst, um die Ausbreitungsverzögerung (t_t bzw. t_{nt} in Abbildung 7) des Echos ergänzt und dann für die künstlichen neuronalen Netze als Eingangswerte verwendet (wie bereits am Ende des Abschnitts „Messaufbau und -ablauf“ angesprochen).

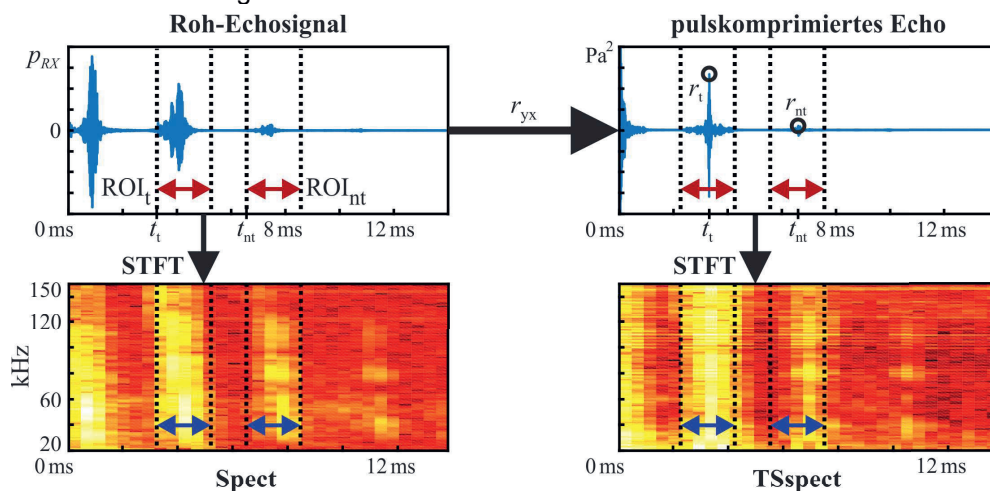


Abb. 7: Vorgehen bei der Berechnung der Features, ausgehend vom Roh-Echosignal mit Kreuzkorrelation (r_{xy}), Kurzzeitfouriertransformation (STFT) und den ROIs für eine Landmarke (ROI_t) sowie kZ (ROI_{nt}). Die Spitzenwerte r_t und r_{nt} werden für die Berechnung von *TStot* benötigt.

Klassifizierungsergebnisse

Nachfolgend werden Klassifizierungsergebnisse aufgeführt und diskutiert, welche mit KNNs erzielt wurden. Für jeden Punkt auf den Abszissen der Abbildungen 8 bis 12 wurden jeweils 20 KNNs trainiert und deren Performanz statistisch anhand ihrer Korrektklassifikationsraten ausgewertet. Es handelt sich dabei um neue KNNs im Vergleich zur Veröffentlichung [13] der Autoren, wobei die Implementierung noch optimiert worden ist.

Es erfolgt eine statistische Auswertung, da durch die zufällige Initialisierung der KNNs zu Beginn eines Trainingsdurchlaufs verschiedene lokale Optima erreicht werden können. Das KNN mit dem besten Ergebnis würde dann für den praktischen Einsatz herangezogen werden. Daher sind die besten erzielten Ergebnisse auch in den Graphen eingetragen. Die Mittelwerte und Standardabweichungen geben eine Orientierung bezüglich des Trainingsaufwands, der nötig ist, um eine gute Lösung zu erzielen. So wird bei einem großen Abstand zwischen bestem Wert und Mittelwert mit Standardabweichung eine höhere Anzahl an Trainingsdurchläufen nötig sein, um ein KNN mit guter Performanz zu finden.

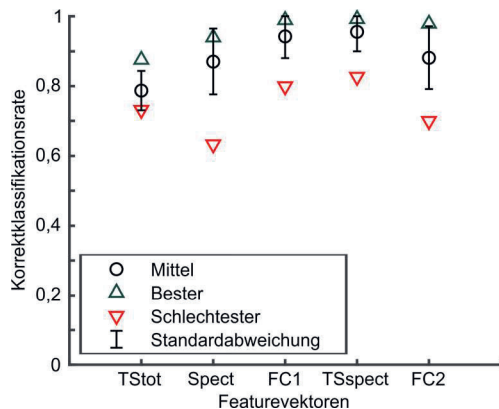


Abb. 8: Korrektklassifikationsrate für breitbandige Downchirp-Anregung, 1,25 MS/s und ohne zusätzliches Rauschen.

Anhand der Abbildungen 8 und 9 ist zu erkennen, dass mit allen Features eine korrekte Klassifizierung der Ziele zu einem hohen Grad ($> 80\%$) möglich ist. Die schlechtesten Ergebnisse werden mit TStot erzielt, wobei hier der geringste Berechnungsaufwand erforderlich ist, da kein Spektrogramm berechnet werden muss. Für Systeme mit begrenzten Ressourcen kann der Einsatz von TStot also sehr wohl sinnvoll sein. Die besten Ergebnisse sind erkennbar für die Spektrogramm-basierten Features, wobei kaum Unterschiede zwischen den besten Ergebnissen für FC1, TSspect und FC2 zu sehen sind.

Interessant ist die Verbesserung von FC1 gegenüber Spect. Anscheinend können mehr relevante Informationen aus den Rohdaten (Spect) extrahiert werden, wenn TStot direkt mit vorgegeben wird, was einer Kombination aus Feature Engineering und Feature Learning entspricht.

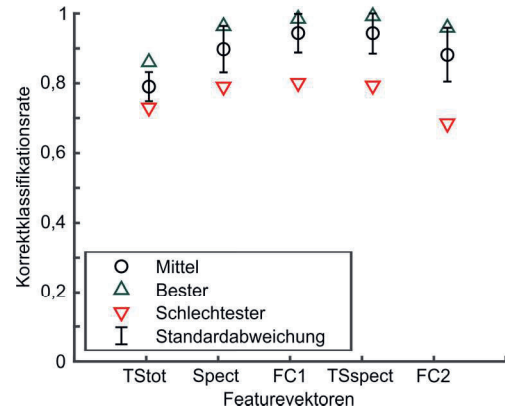


Abb. 9: Korrektklassifikationsrate für schmalbandige Downchirp-Anregung, 1,25 MS/s und ohne zusätzliches Rauschen.

Die Abbildungen 10 bis 12 verdeutlichen den Einfluss von zusätzlichem weißem Rauschen, welches zu den Rohsignalen hinzuaddiert wurde. Eine korrekte Klassifikation von über 90 % der Testdaten ist mit den besten KNNs bei 70 dB re 20 μ Pa zusätzlichem Rauschen bei einer Abtastrate von 1,25 MS/s noch möglich (Abbildungen 10 und 11). Die Fähigkeit von KNNs mit verrauschten Signalen umzugehen, wird demnach als bestätigt angesehen.

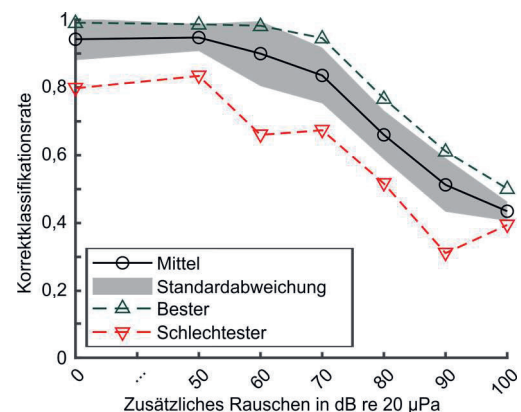


Abb. 10: Korrektklassifikationsrate für FC1 bei breitbandiger Downchirp-Anregung, 1,25 MS/s und mit zusätzlichem Rauschen.

Außerdem ist in Abbildung 12 die erzielbare Korrektklassifikationsrate für schmalbandige Anregungssignale bei einer herabgesetzten Abtastrate von 192 kS/s zu sehen, welche auf Daten basiert, die durch Downsampling der

echten Messdaten bei 1,25 MS/s berechnet wurden. Es ist deutlich zu erkennen, dass auch hier noch über 90 % der Testdaten korrekt klassifiziert werden können bei Rauschschalldrücken von bis zu 60 dB re 20 μ Pa.

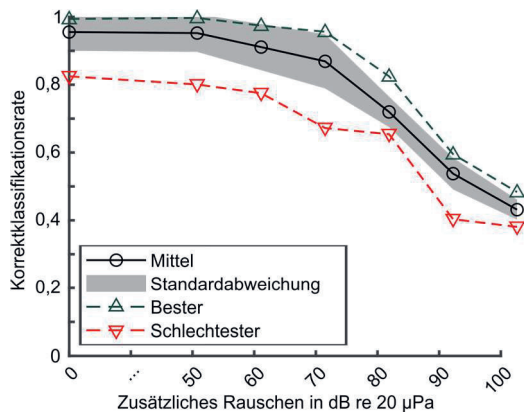


Abb. 11: Korrektclassifikationsrate für TSspect bei breitbandiger Downchirp-Anregung, 1,25 MS/s und mit zusätzlichem Rauschen.

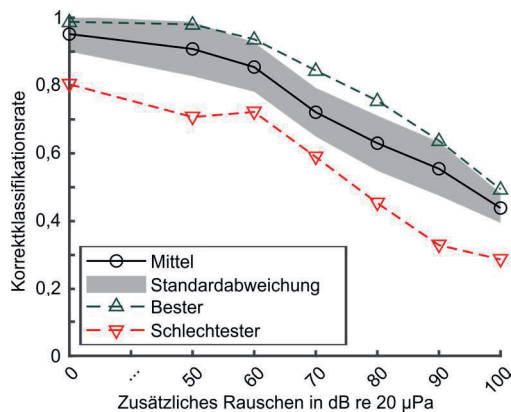


Abb. 12: Korrektclassifikationsrate für FC1 bei schmalbandiger Downchirp-Anregung, 192 kS/s und mit zusätzlichem Rauschen.

Zusammenfassung und Ausblick

Es wurde eine erfolgreiche Klassifizierung von akustischen Landmarken verschiedener Ziele mit den entwickelten Features und KNNs demonstriert. Dabei wurde auch der Einfluss von Anregungssignalen und Rauschen untersucht, wobei sich gezeigt hat, dass sowohl der Einsatz von schmalbandigen wie auch breitbandigen Signalen zu vergleichbaren Ergebnissen führt. Außerdem hat sich die Annahme bestätigt, dass KNNs auch mit verrauschten Signalen die gestellte Aufgabe bewältigen. Die positiven Ergebnisse in Bezug auf Rauschen, Abtastrate und Signalbandbreite motivieren zudem den zukünftigen Einsatz von günstigen MEMS-Mikrofonen mit handelsüblichen Analog-Digital-Umsetzern in Mikrocontrollern und

möglicherweise auch schmalbandigen piezoelektrischen Ultraschallwandlern als Lautsprecher [31]. In Zukunft soll außerdem ein eingebettetes Gesamtsystem implementiert werden, das mehrere Landmarken simultan erfassen und orten kann. Ergänzend soll auch der Einfluss von weiteren Reflexionen an unbekanntem Objekten untersucht werden, die zusätzlich zu den Landmarken vorhanden sein können.

Literaturnachweis

- [1] L. Kleeman, R. Kuc, Sonar Sensing. In Springer Handbook of Robotics; Siciliano, B.; Khatib, O., Eds.; Springer International Publishing: Cham; 491–519 (2016);
- [2] J. Steckel, H. Peremans, BatSLAM: Simultaneous localization and mapping using biomimetic sonar. *PLoS one*, 8, e54076 (2013); doi: 10.1371/journal.pone.0054076
- [3] H.R. Everett, Sensors for mobile robots: Theory and application / H.R. Everett; A. K. Peters: Wellesley, Mass. (1995)
- [4] R.J. Przybyla, H.Y. Tang, Guedes, A.; Shelton, S.E.; Horsley, D.A.; Boser, B.E. 3D Ultrasonic Rangefinder on a Chip. *IEEE Journal of Solid-State Circuits*, 50, 320–334 (2015); doi: 10.1109/JSSC.2014.2364975
- [5] D. Vanderelst, J. Steckel, A. Boen, H. Peremans, M.W. Holderied, Place recognition using batlike sonar. *eLife*, 5 (2016); doi: 10.7554/eLife.14188
- [6] S. Thrun, W. Burgard, D. Fox, Probabilistic robotics; Intelligent robotics and autonomous agents series, MIT Press: Cambridge, Massachusetts and London, England (2006);
- [7] P. Rashidi, A. Mihailidis, A Survey on Ambient-Assisted Living Tools for Older Adults. *IEEE Journal of Biomedical and Health Informatics*, 17, 579–590 (2013); doi: 10.1109/JBHI.2012.2234129
- [8] T. Dahl, M. Boulos, Robots in Health and Social Care: A Complementary Technology to Home Care and Telehealthcare? *Robotics*, 3, 1–21 (2014); doi: 10.3390/robotics3010001
- [9] I. Eliakim, Z. Cohen, G. Kosa, Y. Yovel, A fully autonomous terrestrial bat-like acoustic robot. *PLoS computational biology*, 14, e1006406 (2018); doi: 10.1371/journal.pcbi.1006406
- [10] B. Barshan, B. Ayrulu, S.W. Utete, Neural Network-Based Target Differentiation Using Sonar for Robotics Applications. *IEEE Transactions on Robotics and Automation*, 16, 435–442 (2000); doi: 10.1109/70.864239
- [11] B. Ayrulu, B. Barshan, Neural networks for improved target differentiation and localization with sonar. *Neural Networks*, 14, 355–373 (2001); doi: 10.1016/S0893-6080(01)00017-X
- [12] M. Dmitrieva, M. Valdenegro-Toro, K. Brown, G. Heald, D. Lane, Object classification with convolution neural network based on the time-frequency representation of their echo. In *Proceedings of MLSP2017*; Ueda, N., Ed.; IEEE: Piscataway, NJ; 1–6 (2017); doi: 10.1109/MLSP.2017.8168134

- [13] P.K. Kroh, R. Simon, S.J. Rupitsch, Classification of Sonar Targets in Air: A Neural Network Approach. *Sensors* (Basel, Switzerland), 19 (2019); doi: 10.3390/s19051176
- [14] R. Simon, M.W. Holderied, O. von Helversen, Size discrimination of hollow hemispheres by echolocation in a nectar feeding bat. *The Journal of experimental biology*, 209, 3599–3609 (2006); doi: 10.1242/jeb.02398
- [15] I. Goodfellow, Y. Bengio, A. Courville, Deep learning; Adaptive computation and machine learning, The MIT Press: Cambridge Massachusetts (2016);
- [16] C.M. Bishop, Pattern recognition and machine learning; Information science and statistics, Springer: New York (2006);
- [17] S.S. Haykin, Neural networks and learning machines, 3rd ed. ed.; Prentice Hall: New York (2009);
- [18] G. Dougherty, Pattern Recognition and Classification; Springer New York: New York, NY (2013); doi: 10.1007/978-1-4614-5323-9
- [19] C.E. Metz, Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283–298 (1978); doi: 10.1016/S0001-2998(78)80014-2
- [20] T. Fawcett, An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874 (2006); doi: 10.1016/j.patrec.2005.10.010
- [21] M.I. Skolnik, Radar handbook, 3rd ed. ed.; McGraw-Hill: New York (2008);
- [22] D.A. Abraham, Signal Processing. In *Applied Underwater Acoustics*; Elsevier; 743–807 (2017); doi: 10.1016/B978-0-12-811240-3.00011-4
- [23] J.P. Marage, Y. Mori, Sonar and underwater acoustics; Iste Wiley: London and Hoboken, N.J (2010); doi: 10.1002/9781118600580
- [24] D.A. Kiefer, M. Fink, S.J. Rupitsch, Simultaneous Ultrasonic Measurement of Thickness and Speed of Sound in Elastic Plates Using Coded Excitation Signals. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64, 1744–1757 (2017); doi: 10.1109/TUFFC.2017.2746900
- [25] R.P. Hodges, Underwater acoustics: Analysis, design, and performance of sonar; Wiley: Hoboken, NJ (2010); doi: 10.1002/9780470665244
- [26] T. Afouras, J.S. Chung, A. Senior, O. Vinyals, A. Zisserman, Deep Audio-visual Speech Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018); doi: 10.1109/TPAMI.2018.2889052
- [27] J.H. Connolly, E.A. Edmonds, J.J. Guzy, S.R. Johnson, A. Woodcock, Automatic speech recognition based on spectrogram reading. *International Journal of Man-Machine Studies*, 24, 611–621 (1986); doi: 10.1016/S0020-7373(86)80012-8
- [28] S. Ganapathy, Multivariate Autoregressive Spectrogram Modeling for Noisy Speech Recognition. *IEEE Signal Processing Letters*, 24, 1373–1377 (2017); doi: 10.1109/LSP.2017.2724561
- [29] J.F. Gemmeke, T. Virtanen, A. Hurmalainen, Exemplar-Based Sparse Representations for Noise Robust Automatic Speech Recognition. *IEEE Transactions on Audio, Speech, and Lan-*
guage Processing, 19, 2067–2080 (2011); doi: 10.1109/TASL.2011.2112350
- [30] V. Zue, L. Lamel, An expert spectrogram reader: A knowledge-based approach to speech recognition. *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical and Electronics Engineers, 1197–1200 (1986); doi: 10.1109/ICASSP.1986.1168798
- [31] S.J. Rupitsch, Piezoelectric Sensors and Actuators – Fundamentals and Applications; *Topics in Mining, Metallurgy and Materials Engineering*, Springer Berlin Heidelberg: Berlin, Heidelberg (2019); doi: 10.1007/978-3-662-57534-5