

Abschlussbericht

Aktenzeichen: 36-3400.7/100
Projekt: EmbeddedNeuroVision
Projektpartner: FZI, HSA, Inferics



Anzahl beteiligter Unternehmen: 2 (HSA, Inferics)

Darin enthaltene Anzahl beteiligter KMU: 2 (HSA, Inferics)



**FZI Forschungszentrum
Informatik
(FZI)**

Haid-und-Neu-Str. 10-14
76131 Karlsruhe

Victor Pazmino
pazmino@fzi.de
+49 (0)721 9654 190



**HS Analysis GmbH
(HSA)**

Haid-und-Neu-Straße 7
76131 Karlsruhe

Sergey Biniaminov
[sergey.biniaminov@hs-
analysis.com](mailto:sergey.biniaminov@hs-analysis.com)
+49 (0)7247 93429980



**Inferics GmbH
(Inferics)**

Benzstr. 3
76185 Karlsruhe

Dr. Norbert Link
norbert.link@inferics.com
+49(0)721 721 3528 732

Der Bericht wird auf der Projektwebseite:

<https://www.embeddedneurovision.de/>

veröffentlicht und wird zusätzlich noch von den Partnern unter folgenden Internetadressen zur Verfügung gestellt:

<https://cloud.inferics.com/index.php/s/aWQMqtJg2Z9dHXK>

<https://www.hs-analysis.com/embnv/>

Inhaltsverzeichnis

1.	Projektergebnisse	3
	Projektziel.....	3
	Meilensteine	3
	Abweichung zur ursprünglichen Planung	13
2.	Verbreitung der Ergebnisse	14
3.	Verwertung und Transfer	16
	Wissenstransfer	16
	Transferkonzept zur kommerziellen Verwertung	16

1. Projektergebnisse

Projektziel

Das Ziel des Projekts EmbeddedNeuroVision war der Vergleich von neuromorphen Sensoren und KI-Systemen mit aktuell eingesetzten, nicht-neuromorphen Varianten. Konkret sollte im Kontext von Aktions- bzw. Gestenerkennung („Human Action Recognition“) untersucht werden, wie sich ein System bestehend aus neuromorpher Kamera (auch Event-basierte Kamera genannt) und neuromorpher KI-Verarbeitung (Spiking Neural Network auf neuromorpher Hardware) gegen ein klassisches System aus Bild-basierter Kamera und künstlichem neuronalen Netz schlägt. Zudem sollte untersucht werden, ob sich einzelne Verarbeitungsschritte der beiden Systeme austauschen lassen, also bspw. eine klassische Kamera durch geschickte Konvertierung der Daten in Kombination mit neuromorpher KI-Verarbeitung verwendet werden kann.

Meilensteine

Die Projektarbeit war in sechs Arbeitspakete unterteilt, deren Fortschritt und Abschluss über die in Tabelle 1 aufgelisteten Meilensteine überprüft wurde. Es wurden alle Arbeiten erfolgreich abgeschlossen. Eine kurze Beschreibung der Arbeiten und Ergebnisse sowie ggf. Abweichungen vom Projektplan folgen für jeden Meilenstein im Anschluss an die Tabelle.

Tabelle 1: Meilensteine inklusive Bearbeitungsstatus.

MS	Beschreibung	Status
M0.1	Projektwebseite und Logo wurde aufgesetzt	abgeschlossen
M1.1	Anforderungen, Schnittstellen und Systemarchitektur definiert	abgeschlossen
M2.1	Integration konventioneller Kamera abgeschlossen	abgeschlossen
M2.2	Konvertierungsalgorithmus implementiert	abgeschlossen
M2.3	Event-basierte Kamera integriert	abgeschlossen
M3.1	Künstliche neuronale Netze ausgewählt	abgeschlossen
M3.2	Trainierte KNN vorhanden	abgeschlossen

M3.3	KNNs wurden basierend auf Event-basierten Daten neu trainiert	abgeschlossen
M4.1	Spiking Neural Network ausgewählt	abgeschlossen
M4.2	Trainiertes SNN vorhanden	abgeschlossen
M4.3	Evaluation des SNN abgeschlossen	abgeschlossen
M5.1	Evaluation abgeschlossen, Projektergebnisse liegen vor	abgeschlossen

M0.1 Projektwebseite und Logo: Zu Beginn des Projekts wurde ein passendes Logo entworfen, welches die Hauptaspekte „Vision“ und „KI“ kombiniert und so die Themen des Projekts schnell erfassbar machen soll. Zudem wurde eine Webseite aufgesetzt, welche die Projektziele, -motivation und -ergebnisse für die Öffentlichkeit aufbereitet:

<https://www.embeddedneurovision.de/>

M1.1 Anforderungen, Schnittstellen und Systemarchitektur: Im Rahmen eines gemeinsamen Brainstormings wird die Systemarchitektur nach dem Design-Thinking-Ansatz entworfen: Die grundlegenden Elemente bilden die eventbasierte Kamera zur „Bild“-Erfassung und die anschließende Verarbeitung auf einem Computer. Beide Teile werden als einzelne Module erfasst und sollen autark agieren. Im Gegensatz zu allgemein bekannten Frame-basierten Kameras, die zeitlich diskrete, näherungsweise synchrone Abbilder des gesamten Sensors (Frames) abbilden, nehmen Event-basierte Kameras Veränderungen im Ladungsstrom einzelner Pixel des Sensors wahr und geben somit als Signal eine Folge asynchroner Ereignisse aus. Der resultierende Datenstrom beinhaltet somit Informationen über den Ort und die Stärke einer Veränderung, unabhängig vom Zustand der restlichen Pixel. Daraus ergeben sich andere Anforderungen für den Entwurf des Gesamtsystems: in Bezug auf die Echtzeit des Systems bleibt weiterhin die Abtastrate relevant, die hierbei in Mega-Events pro Sekunde (MEPS) gemessen wird und somit um mehrere Größenordnungen höher ist, als bei Frame-Kameras. Gleichzeitig ist sie immer noch zeitlich begrenzt, sodass zwischen einer Frequenz- und einer Zeitkodierung (Rate-Coding vs. Temporal Coding) entschieden werden muss. Beim Rate-Coding wird die Anzahl der Impulse innerhalb eines Abtastzeitraums erfasst, sodass eine hohe Anzahl an Impulsen (Spikes) für ein stärkeres Signal sorgt. Beim Temporal Coding wird ein einzelner Spike innerhalb des Abtastzeitraums erfasst, dessen Zeitpunkt des Eintritts relativ zum Abtastzeitraum die

Stärke des Signals bestimmt. Das Rate Coding ermöglicht die Konvertierung konventioneller ANNs, während Temporal Coding kleinere, effizientere Netze erlaubt. Im Hinblick auf das formulierte Ziel, ein energiesparendes Produkt zu entwerfen, ist es von Vorteil, in diesem Projekt das Temporal Encoding zu verwenden. Auf der Hardware-Ebene kann hingegen mit ähnlichen Anforderungen wie bei Frame-Kameras gearbeitet werden. Hier wird als Anforderung ein standardisiertes USB-Interface verwendet, für eine spätere, tiefergehende Integration wird ein MIPI-CSI2 zur Leiterplattenmontage vorgesehen, sodass als Schnittstelle zur SNN-verarbeitenden Hardware ein normales USB-Typ A Interface verwendet werden kann.

M2.1 Integration konventioneller RGB-d-Kamera: Die Hardware-Grundlage bildet das Produkt PatronuSens Professional der Inferics GmbH, welche mittels einer omnidirektionalen Stereo-Kamera sowohl RGB-Bilddaten aufnimmt, als auch daraus 3D-Daten des beobachteten Halbraumes berechnet, so dass sie Ströme von RGB-3D-Daten (zu jedem Pixel auch die x,y,z-Koordinaten) liefert sowie mittels der eingebauten KI auch einen Strom von 3D-Koordinaten der Keypoints (Gelenke, Augen, Ohren, Nase) aller im Raum befindlichen Personen. Entsprechend dem Vorgehen bei der Verarbeitung von Daten aus Event-Kameras, bei denen zur Verarbeitung aus dem Eventdatenstrom Frames mit Pixelaktivierungen entsprechend der gelieferten Events erzeugt werden, werden aus aufeinanderfolgenden Frames der konventionellen Kamera sogenannte Eventframes erzeugt, in denen die Aktivierung sich überschwellig ändernder Pixel dargestellt ist, wobei dies RGB-Eventframes, 3D-Eventframes oder Keypoint-Eventframes sein können. Die so geschaffene „Eventframe-Kamera“ benutzt somit die Bildströme ihrer integrierten konventionellen PatronuSens-Framekamera, um daraus einen Eventframe-Datenstrom zu erzeugen, der qualitativ den Daten entspricht, wie sie aus einer Event-Kamera verarbeitet werden.

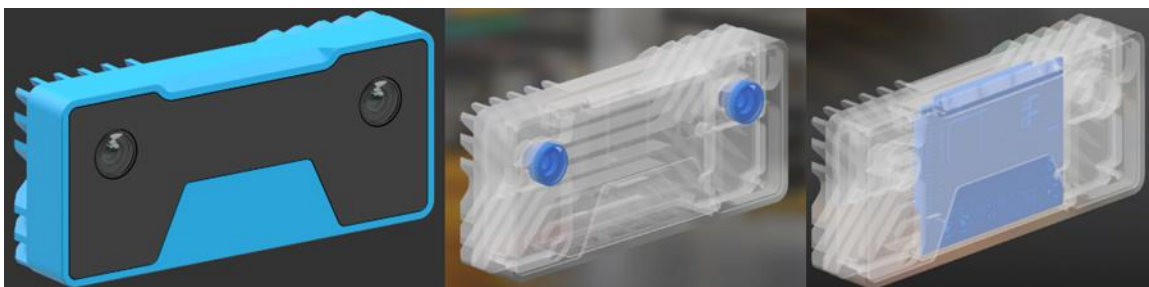


Abbildung 1: Hardware der EmbeddedNeuroVision Eventframe-Kamera

Dabei wird die Eventframe-Kamera als Softwaremodul realisiert, welches die jeweils aktuellen Eventframedaten aus den Daten der jeweils letzten beiden Frames der eingebauten Kamera erzeugt und über eine Softwareschnittstelle dem Auswertemodul im Eventframeformat zur Verfügung stellt, wie in nachfolgender Grafik dargestellt.

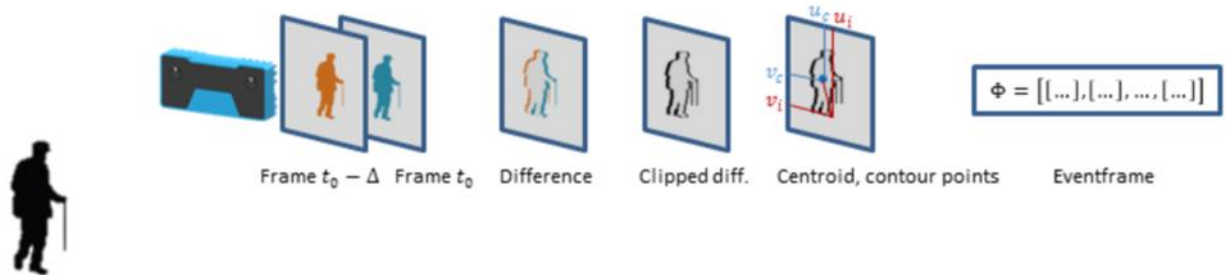


Abbildung 2: Architektur der Eventframe-Kamera

M2.2 Konvertierungsalgorithmus Frame-Kamera-Datenströmen zu Event-Kamera-Datenströmen und Eventframe-Datenströmen: Steht nur eine Framekamera zur Verfügung kann der zeitliche Zusammenhang in Anlehnung an die Eventframedarstellung erfolgen, um sich den Datenströmen der Event-Kamera anzunähern. Dies kann durch eine Framekamera angenähert werden, wenn die Farbwertänderungen zwischen zwei Frames ausgewertet werden. Eine derartige Emulation von Event-Kameras wurde bereits in der Literatur vorgeschlagen, wo für jedes Pixel zweier aufeinander folgender Frames eine lin-log-Funktion (Annäherung des Logarithmus) der Intensität I (gebildet aus den RGB-Werten $I=R+G+B$) oder des Farbwertes $h = b/(r+g)$ berechnet wird. Die Differenz dieser Werte zwischen zwei Frames wird auf eine betragsmäßig überschwellige Änderung untersucht und im positiven Fall dem Pixel der Wert ON oder OFF (je nach Vorzeichen der Differenz) zugewiesen. Da nun zu einem Zeitpunkt mehrere Pixel gleichzeitig ON/OFF-Werte zugewiesen bekommen haben und viele Auswerteverfahren gleiche Zeitstempel nicht verarbeiten können, wird das Zeitintervall zwischen zwei Frames entsprechend der Anzahl der simultanen Pixel linear interpoliert und die Interpolationswerte als individuelle Zeitstempel zugewiesen. Diese etablierte Art der Emulation führt zwar zu Datenströmen, welche denen von Event-Kameras entsprechen, die jedoch so von den gängigen Auswerteverfahren durch Faltungsschichten (Convolutions) und Spiking Neural Networks gar nicht verarbeitet werden und bei direkter Verarbeitung mit rekurrenten Netzen keine zufriedenstellenden Ergebnisse zeitigen, wie in den Arbeiten gezeigt wurde. Vielmehr findet als Vorverarbeitung eine Aggregation der Events statt, indem die Event-Koordinaten für die

Erstellung eines „Event-Bildes“ bzw. „Eventframe“ verwendet werden, in welchem Events der entsprechenden Pixelkoordinate zu einem Eintrag an der entsprechenden Position im Bild mit der zugehörigen Polarität führen. Daher wurde im Vorhaben ein Konvertierungsalgorithmus für Datenströme aus einer Frame-Kamera entwickelt, der genau dieser Darstellung entsprechende Eventframe-Datenströme liefert, welche analog zu denen der Event-Kamera ausgewertet werden können.

Bei der Eventframe-Kamera entsteht also eine Menge Y von Pixeln gleichzeitiger Events („Eventframe“), welche die Änderungen im Farbwert über einer Schwelle zwischen zwei Bildframes darstellen: Berechne das Differenzbild aufeinander folgender Frames $I(t)$ und $I(t + \Delta)$ mit einstellbarer Framerate. Bestimme die Menge Y überschwellig veränderter Pixel:

$$[u, v, \Delta r, \Delta g, \Delta b]^T \in Y, \text{ wenn } \sqrt{([r, g, b]_{u,v,t}^T - [r, g, b]_{u,v,t+\Delta}^T)^2} > c$$

Trage in diesem Fall die Werte $\Delta r, \Delta g, \Delta b$ an der entsprechenden Pixelposition u, v ein, ansonsten den Wert 0.

M2.3 Event-basierte Kamera integriert: Als Hardware werden die Development Kits CeleX5_MP der Firma CelePixel Technology Co. LTD sowie Gen 4.0 der Firma Prophesee verwendet. Anzumerken ist, dass die Firma CelePixel im Laufe des Projekts von Omnivision Sensor Solution Co. LTD aufgekauft worden ist, sodass dieses spezielle Kit nun nicht mehr erhältlich ist. Als Basis für die Integration dient die öffentlich zugängliche Dokumentation der Ansteuerung und Dekodierung der Daten, die hier zu finden ist:

<https://github.com/CelePixel/CeleX5-MIPI>.

Die CelePixel-Kamera bietet mehrere Modi zur Datenerfassung, derjenige, welche dem Temporal Encoding (beschrieben in M1.1) entspricht ist der Modus Event-In-Timestamp. Hierbei hat jeder Datenpunkt die Struktur:

Zeile [py], Spalte [px], Eintrittszeitpunkt [μ s], Übertragungszeitpunkt [μ s]

Somit kann das Temporal Encoding aus einer gespeicherten Datei sowie aus dem kontinuierlichen Datenstrom entnommen werden, indem die Koordinaten und der Eintrittszeitpunkt der Ereignisse berücksichtigt werden.

Im Gegensatz dazu werden die Daten der Prophese-Kamera standardmäßig in folgendem Format ausgegeben, welche zusätzlich die Polarität der Events enthält (also ob ein Pixel heller oder dunkler geworden ist):

Spalte [px], Zeile [py], Polarität [binär], Übertragungszeitpunkt [μ s].

Die Polarität wurde im Rahmen des Projekts für eine bessere Vergleichbarkeit jedoch nicht verwendet. Für eine erfolgreiche SNN-Analyse mit den vorliegenden SNN-Frameworks wurde in der Praxis der Full-Picture-Modus verwendet, da eine vollständige Matrix für die Analyse in den nachfolgenden SNN, insbesondere der Anbindung an die neuromorphe Hardware, erforderlich ist. Hierbei werden alle Ereignisse in einem definierten Zeitraum in eine Matrix zusammengefasst mit den Dimensionen der Auflösung der Kamera zusammengefasst und gesammelt als Frame an die Analyse weitergegeben.

M3.1 Entwicklung Künstlicher Rekurrenter Neuronaler Netze zur Aktivitätserkennung aus Framekamera- und Eventframe-Datenströmen: Als Baseline diente die bereits zuvor von Inferics entwickelte Personen-Aktivitätserkennung aus Datenströmen von Framekameras, bei denen zunächst für jeden Frame (bestehend aus einem Stereo-Bildpaar) die 3D-Koordinaten der 17 Körper-Keypoints (Gelenke und die Kopf-Keypoints Augen, Ohren und Nase) mittels eines tiefen Faltungsnetzes extrahiert werden, aus diesen ein 17x3 Posenframe berechnet und Folgen dieser Posenframes mit einem Zeitfaltungsnetz (t-CNN) verarbeitet werden. Diesem Basisverfahren („Pose+t-CNN“) sollte ein neuartiger Ansatz auf Basis rekurrenter Netze gegenübergestellt werden: Zur einheitlichen Verarbeitung zeitlich unterschiedlich langer Vorgänge sollte ein „unkonventionelles“ Echo-State-Network, ESN, (stark verwandt mit Spiking Neural Networks, wie vom Partner FZI eingesetzt), das direkt unterschiedliche Zeitverläufe durch interne Zustände (Echo States) darstellt, ein Embedding der Zeitsignale in Form der Zustandsausprägungen repräsentieren, welche dann als Zustandsfolge von einem rekurrenten Netz des Typs LSTM zur Erkennung verarbeitet werden sollte, wie in unten stehender Grafik dargestellt.

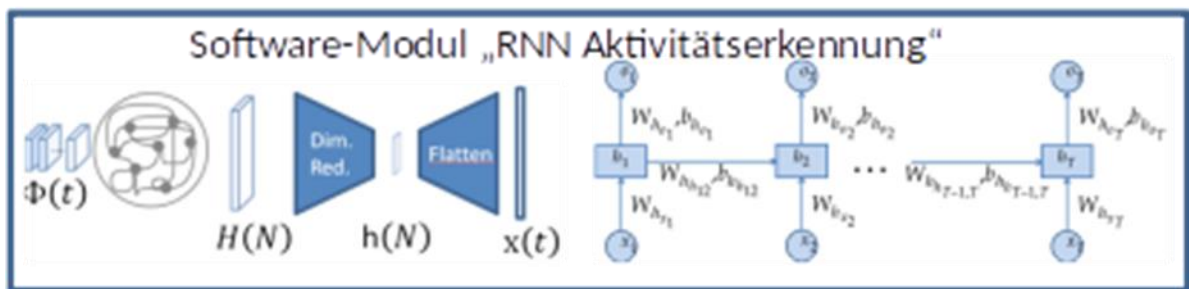


Abbildung 3: Ursprünglicher Ansatz mit ESN (links) und LSTM (rechts)

Vorstudien der ESN mit dem vereinfachten Fall der Erkennung pathologischer Zustände in einkanaligen EKG-Datenströmen zeigten jedoch eine deutlich zu geringe Robustheit der Ergebnisse, so dass dieser Weg nicht weiter verfolgt werden konnte.

In einem zweiten, alternativen, rekurrente-Netze-basierten Ansatz („Pose+RNN“) wurde die Vorverarbeitung durch das ESN ersetzt, indem hierfür stattdessen die Erzeugung von Posenframe-Datenströmen aus dem Baseline-Ansatz verwendet wurde, und eine nachgeschaltete Aktivitätserkennung aus Posenframe-Datenströmen mittels LSTM genutzt wurde, um die Aufgabe zu lösen, was sehr gute Ergebnisse lieferte und als danach Vergleichsmaßstab zum Einsatz von SNNs auf Datenströmen aus der Event-Kamera diente.

M3.2 Training und Evaluierung der Aktivitätserkennung aus Framekamera-Datenströmen: Vor der Erhebung projektspezifischer Daten wurde das „Pose+t-CNN“-Verarbeitungsnetz mit Daten des öffentlich zugänglichen NTU-RGB-d-Datensatzes zur Erkennung ausgewählter Aktivitäten trainiert und evaluiert, welches dabei State-of-the-Art Leistungen aufwies.

M3.3 Training und Evaluierung der Aktivitätserkennung aus Eventframe-Datenströmen: Anstelle des ursprünglich vorgesehenen Netzwerks zur Verarbeitung von Eventframe-Datenströmen, welches sich in den Voruntersuchungen bei der Entwicklung der Architektur als unbefriedigend erwiesen hat, wurde wie in M3.1 ausgeführt „Pose+RNN“ als Alternative entwickelt. Ebenso wurde das „Pose+RNN“-Netz auf Posenframe-Daten trainiert, welche aus dem NTU-RGB-d-Datensatz gewonnen wurden und vor der Leistung dem „Pose+t-CNN“-Verarbeitungsnetz in etwa gleichwertig befunden. Das „Pose+RNN“-Netz wurde dann ausgewählt, um gegen ein SNN anzutreten (welches Daten direkt von der Eventkamera verarbeitet) und entsprechend mit den Daten aus dem im Rahmen des Vorhabens erhobenen Datensatz trainiert.

M4.1 Auswahl der Spiking Neural Networks: Im Rahmen des Projekts wurden zunächst verschiedene Frameworks zur Entwicklung von Spiking Neural Networks (SNN) untersucht, da diese einen direkten Einfluss auf die möglichen SNN-Modelle haben. Erste Tests mit Nengo¹, PyNN² und BindsNET³ wurden auf Grund fehlender Funktionalität oder komplexer Erweiterbarkeit/Integration wieder eingestellt. Letztendlich wurde Norse⁴ in Kombination mit Tonic⁵ als Framework verwendet, da diese ausreichende Funktionalität, schnelle Ergebnisse und eine einfache Integration von Event-Kamera-Daten versprachen. Als SNN-Modelle wurden rekurrente, faltungsbasierte (convolutional) und vollständig vernetzte Varianten gewählt.

M4.2 Training der Spiking Neural Networks: Es wurden mehrere SNN-Modelle auf verschiedenen Datensätzen trainiert. Zunächst wurden synthetische Daten basierend auf dem zu M2.2 entwickelten Konvertierungsalgorithmus verwendet. Später wurde im Rahmen des Projekts ein Datensatz aufgenommen, welcher aus zwölf Klassen und insgesamt 580 Aufnahmen besteht, die jeweils mit einer Frame-Kamera und zwei Event-Kameras parallel aufgenommen wurden. Von diesen zwölf Klassen sind zehn Klassen an den neuromorphen „DVS Gesture“-Datensatz⁶ angelehnt, welche um die Klassen „Hinsetzen“ und „Aufstehen“ ergänzt wurden, um dem Projektkontext Rechnung zu tragen. Dieser Datensatz wurde verwendet um mehrere SNN zu trainieren, wobei alle Modelle jeweils für 3 Klassen (Aufstehen, Hinsetzen, Andere) und 12 Klassen (alle oben genannten) trainiert wurden.

M4.3 Evaluation der Spiking Neural Networks: Die trainierten SNN wurden basierend auf den aufgenommenen Daten v.a. hinsichtlich Vorhersagegenauigkeit und Parameteranzahl evaluiert. Insbesondere die Parameteranzahl sollte nicht zu groß werden, um eine Einsatzfähigkeit in eingebetteten Systemen zu gewährleisten. Als Richtwert wurden hier ~ 5 Mio. Parameter gewählt. Die Evaluation hat gezeigt, dass reine vollständig verdrahtete und rekurrente Modelle keine konkurrenzfähige Vorhersagegenauigkeit liefern konnten, ohne die Zielparameteranzahl stark zu überschreiten. Bessere Ergebnisse wurden mit Faltungsnetzen erzielt, welche eine

¹ <https://www.nengo.ai/>

² <https://neuralensemble.org/PyNN/>

³ <https://github.com/BindsNET/bindsnet>

⁴ <https://github.com/norse/norse>

⁵ <https://github.com/neuromorphs/tonic>

⁶ <https://research.ibm.com/interactive/dvsgesture/>

ähnliche Genauigkeit erzielen, wie zu M3.2 entwickelten KNN. Die besten Ergebnisse wurden mit einer Kombination aus rekurrenten, voll verdrahteten und Faltungsschichten erzielt, welche bei 3 Klassen eine sehr hohe Genauigkeit von ca. 98% bei weniger als 400.000 Parametern erreichen. Eine genaue Auflistung und Gegenüberstellung der verschiedenen Netze und Genauigkeiten ist in Tabelle 2 zu finden.

Tabelle 2: Vergleich von Genauigkeit der Netze auf den aufgenommenen Daten

Neuronales Netz	Genauigkeit bei 3 Klassen	Genauigkeit bei 12 Klassen	Anzahl Parameter
ANN (CNN + LSTM)	96%	57%	~ 6.500.000
SNN Fully Connected	80%	30%	~ 10.000.000
SNN Convolutional	95%	87%	~ 1.000.000
SNN Conv. + Recurrent	98%	92%	~ 380.000

Die SNN wurden zudem soweit möglich auf die neuromorphe Hardware-Plattform BrainChip Akida⁷ portiert und dort hinsichtlich des Durchsatzes und der elektrischen Leistungsaufnahme evaluiert. Ein Vergleich zur Simulation der SNN auf einem GPU-basierten Rechner sowie mit den Werten des KNN-basierten Systems sind in Tabelle 3 zu sehen.

Tabelle 3: Vergleich von el. Leistung und BPS der neuronalen Netze auf verschiedenen Hardware-Systemen

System	Leistungsaufnahme	Bilder pro Sekunde
KNN auf Nvidia Jetson NX	~5-9W	~1
SNN auf GPU-Rechner	~150W	0,5 - 1
SNN auf BrainChip Akida	<1W	40 - 45

M5.1 Evaluation und Demonstration der Projektergebnisse: Zur Demonstration der Projektergebnisse wurde ein System aufgebaut, welches jeweils eine klassische und eine Event-Kamera integriert. Zudem sind an das System ein Nvidia Jetson NX⁸ als klassischer KI-Prozessor sowie ein BrainChip Akida als neuromorpher KI-Prozessor angeschlossen, auf welchen die neuronalen Netze laufen, welche die Partner im Rahmen des Projekts entwickelt haben. Die Videos beider Kameras werden an eine grafische

⁷ <https://brainchipinc.com/akida-neural-processor-soc/>

⁸ <https://www.nvidia.com/de-de/autonomous-machines/embedded-systems/jetson-xavier-nx/>

Benutzerschnittstelle weitergeleitet, sodass für einen Betrachter die Unterschiede zwischen den verschiedenen Kameras sichtbar werden. Zudem werden die Daten der Event-Kamera an den neuromorphen Prozessor weitergeleitet und dort von einem SNN verarbeitet, während die Daten der klassischen Kamera an den Nvidia Jetson NX weitergeleitet werden. Das Konzept des Demonstrators ist in Abbildung 4 dargestellt.



Abbildung 4: Konzeptioneller Aufbau des Demonstrators

Die Vorhersagen der jeweiligen Netze sowie die benötigte Leistungsaufnahme und die aktuell von den Kameras übertragene Datenrate werden ebenfalls parallel grafisch dargestellt. Die grafische Oberfläche ist in Abbildung 5 zu sehen.

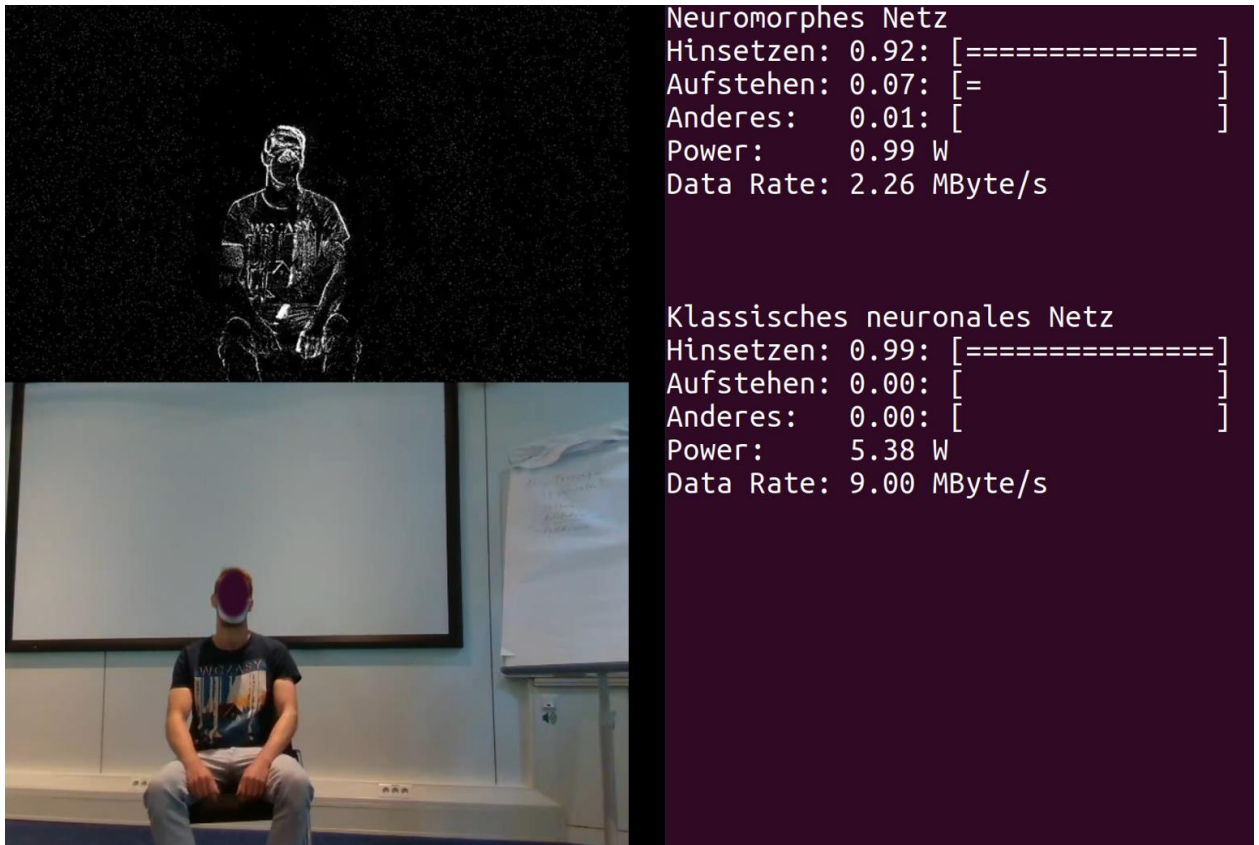


Abbildung 5: Aufnahme des Demonstrator-Interfaces während des Betriebs

Die Demonstration verdeutlicht das Potenzial des neuromorphen Systems bezüglich Energieeffizienz und benötigter Datenmenge, welche beide ungefähr um den Faktor 5 verbessert wurden. Dem gegenüber stehen allerdings die hohen Kosten des neuromorphen Systems und insbesondere der Event-Kamera von mehreren Tausend Euro, welche vor einem erfolgreichen Einsatz im untersuchten Anwendungsfall noch deutlich verringert werden müssen.

Abweichung zur ursprünglichen Planung

Hardwarebeschaffung

Aufgrund der Verzögerung bei der Lieferung der Kamera aus China und einem sehr hohen Aufwand auf der Verwaltungsseite bzw. der defekten Lieferung der Hardware aus Frankreich und erneuten Zulieferungen, kam es zu einer Verzögerung bei der Datenerstellung und anschließenden Integration vom Datenformat in ein für das Projekt notwendiges Format. So konnten unter zeitlichem Druck sogar einige technische Probleme bei der Integration behoben werden, jedoch geschah dies unter höherem

Aufwand als geplant auf der Ressourcen-Seite. Letztendlich konnten wir uns nicht ausreichend Zeit für die Evaluation verschiedener SNN-Architekturen lassen und haben die Untersuchung auf eine Version einer SNN-Architektur und deren Implementierung beschränkt, die die Grundzüge des Vorhabens abbildet, jedoch in weiteren Projekten optimiert werden könnte. Die Erstellung der Applikation und die Integration in eine Software erfolgten ohne Probleme und nach dem zuvor definierten Plan.

SNN-Architektur

Wie bereits unter M3.1 und M3.3 beschrieben, hat sich die ursprünglich geplante Neuronale-Netze-Architektur zur Aktivitätserkennung mittels Rekurrenter Neuronaler Netze unter Nutzung von Echo-State-Networks zur Verarbeitung von Eventframe-Datenströmen nicht bewährt. Daher wurde eine neue Architektur entwickelt, welche LSTM als Rekurrentes Netz zur Erkennung von Vorgängen verwendet, jedoch auf Posen-Datenströmen aufsetzt.

2. Verbreitung der Ergebnisse

Das Projekt und dessen Ergebnisse wurden von allen Partnern auf vielen verschiedenen Kanälen der Öffentlichkeit präsentiert, darunter Konferenzen, Festivals, Vorträge, Webseiten, soziale Medien und weitere Veranstaltungen. Nachfolgend eine detaillierte Aufzählung:

- Präsentation des laufenden Projekts auf dem Wissenschaftsfestival „Effekte“, was von der Stadt Karlsruhe und dem Karlsruher Institut für Technologie im Sommer 2021 für die allgemeine Öffentlichkeit veranstaltet wurde. <https://www.effekte-karlsruhe.de/festival-2021/programm-archiv/>
- Im Arbeitskreis Künstliche Intelligenz der Industrie und Handelskammer (IHK) Karlsruhe wurde das laufende Projekt vorgestellt. Im Vordergrund stand hierbei die Zusammenarbeit von FZI, der Inferics GmbH und der HS Analysis GmbH als Experten für SNNs.
- Bei der Jahresversammlung der IHK Karlsruhe in den Räumlichkeiten der HS Analysis GmbH konnte HS Analysis den Fortschritt im Umgang mit der Event-Kamera und den SNNs praktisch live präsentieren. Teilnehmer dieser

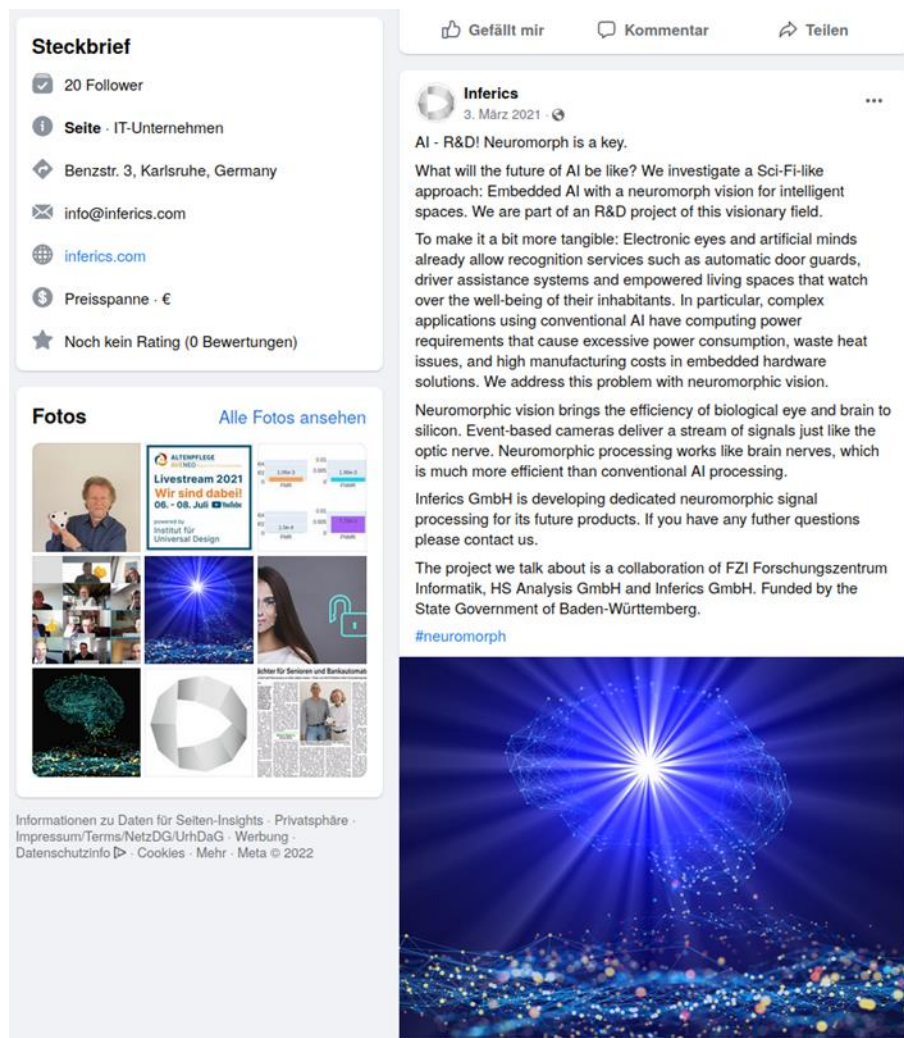
Diskussionsrunde waren unter anderem das Unternehmen IONOS, sowie das Ministerium für Forschung und Wirtschaft des Landes Baden-Württemberg.

- Vortrag am 29.09.2021 beim KI4KMU-Track der Konferenz INFORMATIK2021, veranstaltet von der Gesellschaft für Informatik, in Kombination mit einer Publikation.
- Präsentation im Tollhaus Karlsruhe im November 06.10.2021 KIT Science Week.
- Beiträge auf Facebook

(<https://www.facebook.com/people/Inferics/100057143332472/>)

und LinkedIn

(<https://de.linkedin.com/company/inferics>).



Steckbrief

- 20 Follower
- Seite · IT-Unternehmen
- Benzstr. 3, Karlsruhe, Germany
- info@inferics.com
- inferics.com
- Preisspanne · €
- Noch kein Rating (0 Bewertungen)

Fotos [Alle Fotos ansehen](#)

Informationen zu Daten für Seiten-Insights · Privatsphäre · Impressum/Terms/NetzDG/UrDaG · Werbung · Datenschutzinfo · Cookies · Mehr · Meta © 2022

Inferics
3. März 2021 · 🌐

AI - R&D! Neuromorph is a key.

What will the future of AI be like? We investigate a Sci-Fi-like approach: Embedded AI with a neuromorph vision for intelligent spaces. We are part of an R&D project of this visionary field.

To make it a bit more tangible: Electronic eyes and artificial minds already allow recognition services such as automatic door guards, driver assistance systems and empowered living spaces that watch over the well-being of their inhabitants. In particular, complex applications using conventional AI have computing power requirements that cause excessive power consumption, waste heat issues, and high manufacturing costs in embedded hardware solutions. We address this problem with neuromorphic vision.

Neuromorphic vision brings the efficiency of biological eye and brain to silicon. Event-based cameras deliver a stream of signals just like the optic nerve. Neuromorphic processing works like brain nerves, which is much more efficient than conventional AI processing.

Inferics GmbH is developing dedicated neuromorphic signal processing for its future products. If you have any further questions please contact us.

The project we talk about is a collaboration of FZI Forschungszentrum Informatik, HS Analysis GmbH and Inferics GmbH. Funded by the State Government of Baden-Württemberg.

#neuromorph

- Beitrag auf <https://www.inferics.com/news/>
- Präsentation vor Partnern aus Wissenschaft, Industrie und Politik beim FZI OpenHouse 2022 am 24.02.2022:

<https://www.fzi.de/veranstaltungen/fzi-open-house/> - 12:15 & 13:15 Uhr |

Neuromorphe Systeme und Sensoren

3. Verwertung und Transfer

Wissenstransfer

Die Zusammenarbeit im Projekt war durch einen starken Wissenstransfer zwischen den Partnern geprägt, da jeder Partner eine eigene Kernkompetenz in das Projekt einbringt. HSA fokussierte sich auf die allgemeine Analyse von Daten mit verschiedenen „Deep Learning“-Architekturen. So ist es möglich durch die Zusammenarbeit im Konsortium mit Inferics die Daten mit neuen „Deep Learning“-Technologien auf den Anwendungsfall zu erforschen und anzuwenden. Der Fokus von Inferics lag auf Sensoren zur Erkennung von Aktivitäten, welche batteriebetriebenen über Wochen interaktive Assistenz für Menschen ermöglichen sollen. Inferics stellte die Anwendungsreferenz und entwickelte simulierte neuromorphe Strukturen, welche den „echten“ neuromorphen Systemen der Partner HSA und FZI gegenübergestellt werden. Dadurch kam es zu einem intensiven Austausch von Erfahrungen über Fähigkeiten und Einschränkungen sowie über die notwendigen Aufwände und Einsparungen der konkurrierenden Ansätze. Das FZI brachte sein umfangreiches Wissen in Bezug auf neuromorphe Algorithmen und Sensorik ein, um beide Partner bei der Adaption ihrer Technologien auf ein gemeinsames neuromorphes System zu unterstützen.

Aus dem gegenseitigen Wissensaustausch resultierte ein gemeinsames Verständnis der eingesetzten Technologien, welches es den Projektpartnern ermöglichte einen aussagekräftigen Demonstrator aufzubauen und das Projekt erfolgreich abzuschließen.

Transferkonzept zur kommerziellen Verwertung

HS Analysis

HS Analysis hat großes Interesse seitens unserer bestehenden Kunden sowie einiger möglichen Neukunden erfahren, SNNs in kommerziellen Produkten anzuwenden,

insbesondere im Bereich der Prozessüberwachung. Interesse besteht hierbei vor allem in den reduzierten Energieaufwand und dem erhöhten Datenschutz Event-basierter Kameras. Hiermit lässt sich eine datenschutzkonforme Prozessüberwachung auch in der Öffentlichkeit zugänglichen Bereich ermöglichen, sowie bietet der verringerte Strombedarf die Möglichkeit, On-Edge Geräte mittels Power-over-Ethernet (PoE) zu betreiben, was ein vereinfachtes Deployment ermöglicht.

Unseres bisheriges Kernprodukt, das HSA KIT, welches ein Werkzeugkasten verschiedener angepasster KI-Analysen beinhaltet, können wir zudem durch die im Projekt gewonnen SNN-Kenntnisse erweitern. Dadurch sind nun erweiterte, minimalistische Zeitreihenanalysen möglich, für die wir bereits mit dem Prozess der Angebots- und Bestellungsvergabe begonnen haben.

Inferics

Ob Patienten-Monitoring in Krankenhäusern oder Notfall- und Aktivitätserkennung im betreuten Wohnen oder im Pflegeheim - die Übernahme pflegerischer Routineaufgaben begegnet dem vielzitierten Pflegenotstand und hat daher enormes wirtschaftliches Potential, welches die Inferics GmbH durch entsprechende Vorrichtungen zur Erfassung und Meldung erschließt. Um die Privatsphäre der Pflegeempfangenden zu gewährleisten, erfolgt die Verarbeitung im Sensor selbst, so dass nur Meldungen den Sensor verlassen. Dies bedeutet aber auch, komplexe Erkennungsleistungen in Sensoren mit kleinem Formfaktor durchführen zu müssen. Mit konventioneller Kamera- und Neuronaler-Netz-Technik führt dies zu starken Einschränkungen. Im Projekt wurde sichtbar, dass diese mittels Event-Kameras und Verarbeitung durch wesentlich leichtgewichtiger SNN (insbesondere mit Verfügbarkeit dezidiert hierfür effizienter Hardware) überwunden werden können. Derzeit spricht jedoch der sehr hohe Preis für die Hardware (ein Vielfaches der erzielbaren Gesamtsystemkosten) deutlich gegen eine Marktfähigkeit derartiger Lösungen. Für Inferics lohnt es jedoch, die Preisentwicklung zu beobachten, um aufgrund der im Vorhaben gemachten Erfahrungen rechtzeitig auf die neue Technologie umsteigen zu können. Als Zwischenergebnis des Vorhabens wurde erkannt, dass rekurrente Netze die bisher eingesetzten Zeitfaltungsnetze durch geringeren Ressourcenbedarf vorteilhaft ersetzen können, so dass auf einem Sensor mehrere Erkennungsleistungen gleichzeitig möglich sind, was einen deutlichen Marktvorteil

ergeben kann. Als noch ressourcenschonender haben sich im Projekt die SNN erwiesen, deren Anwendung auf Eventframe-Daten (erzeugt aus Framedaten einer konventionellen Frame-Kamera) den geschilderten Effekt vervielfachen kann. Letzteres könnte einen wirklichen preislichen Durchbruch ergeben, der das Marktpotential noch wesentlich erhöht.