

## **Sachbericht zum Verwendungsnachweis**

### **Teil I – Kurzbericht**

**WIR!-Projekt: Rear-View-Camera for Trains  
(FKZ 03WIR1219A)**

**Zuwendungs-  
empfänger: Technische Universität Chemnitz**

**Laufzeit: 01/2023 – 12/2024**

## 1 Ursprüngliche Aufgabenstellung

Die wesentliche Aufgabenstellung im Projekt ReViCaT lag darin, Videodaten über das im Zusammenhang mit der Einführung der digitalen automatischen Kupplung (DAK) geplante Bussystem zu übertragen, um ein Rückfahrkamerasystem für den Güterzug zu ermöglichen. Dazu sollte auch die Detektionen von Personen im Gefahrenbereich durch intelligente Bildverarbeitung möglich sein. Weiterhin sollte der Abstand zum anzukoppelnden Wagen ermittelt werden, um das Rangieren für den Triebfahrzeugführer zu vereinfachen. Da im Vorfeld nicht absehbar war, ob die Videodatenübertragung vom letzten Wagen zum Führerstand mit einer Latenz behaftet sein wird, war eine Geschwindigkeitsmessung vorgesehen. Mit der Geschwindigkeit sollte ein digitales Zoom die Latenz kompensieren. Mit diesen Punkten zur Aufgabenstellung sollten Algorithmen entwickelt werden, um mit diesen zum Projektende einen Demonstrator zu realisieren und in der realen Einsatzumgebung zu erproben.

## 2 Ablauf des Vorhabens

Zu Beginn des Projekts fand ein Kick-Off-Meeting statt, um detaillierte Punkte zur Organisation abzuklären. Die Arbeitspakete wurden im Wesentlichen in der geplanten Reihenfolge umgesetzt. Der geplante Workshop zu Projektbeginn für die zusätzlichen Anforderungen seitens der Bahnexperten fand erst 09/2023 statt. Die ersten Aufnahmen für die Entwicklung der Algorithmen konnten erst 06/2023 gemacht werden. Damit haben sich die Arbeitspakete zeitlich etwas verschoben, doch ein wesentlicher Verzug fand zunächst nicht statt. Nachdem im August 2024 der Gesamtdemonstrator einsatzbereit war, fanden im Anschluss die ersten Testfahrten statt (Projektmonat 20). Da diese erste Erprobung in der realen Einsatzumgebung Probleme aufzeigte, waren Hardwareanpassungen (andere Kameras) notwendig und damit auch Softwareanpassungen an den Algorithmen. Bei der Beschaffung der neuen Hardware kam es zu erheblichen Verzögerungen, womit die weitere Anpassung der Algorithmen an die neue Hardware und damit zusammen der gesamte Projektverlauf ins Stocken geriet. Der abschließend erfolgreiche Test konnte zum Ende der Projektlaufzeit nicht stattfinden und somit der Demonstrator auch nicht fertiggestellt werden. Eine beantragte kostenneutrale Projektverlängerung wurde nicht bewilligt.

## 3 Zusammenfassung der wesentlichen Ergebnisse

Im Projekt konnte gezeigt werden, dass der Einsatz eines Rückfahrkamerasystems im Güterverkehr prinzipiell möglich ist. Eine Personendetektion konnte erfolgreich umgesetzt werden. Der Abstand konnte mit einer Genauigkeit von 0,14 m im Nahbereich ermittelt werden. Für den Triebfahrzeugführer war das nach eigenen Aussagen eine gute Hilfe. Eine Latenzkompensation war nicht notwendig, weswegen kein digitales Zoom verwendet wurde. Die Funktionen wurden erfolgreich im Laborumfeld umgesetzt, die Portierung auf das Embedded System konnte zum Projektende nicht fertiggestellt werden. Ein echtzeitfähiger Demonstrator lag somit zum Projektende nicht vor.

Die Nachteile eines Rückfahrkamerasystems bestehen darin, dass entweder die Kamera bei jedem Rangievorgang manuell angebracht werden muss oder für jeden Güterwagen zwei Kamerasysteme notwendig wären. Bei zwei Kamerasystemen entstünden erhebliche Kosten. Zusätzlich besteht das Risiko von Verschmutzung, Abnutzung und besonders von Vandalismus. Daher äußerten Bahnexperten Zweifel bzgl. der praktischen Umsetzbarkeit und Einsatzfähigkeit eines solchen Systems.

## **Sachbericht zum Verwendungsnachweis**

### **Teil II – Ausführlicher Bericht**

**WIR!-Projekt: Rear-View-Camera for Trains  
(FKZ 03WIR1219A)**

**Zuwendungs-  
empfänger: Technische Universität Chemnitz**

**Laufzeit: 01/2023 – 12/2024**

## 1 Durchgeführte Arbeiten

- Die ersten Testaufnahmen fanden am 06.06.23 auf dem Güterbahnhof Minden statt, indem mehrere Kameras an dem Ende eines H-Wagens befestigt wurden. Dieser Wagen wurde dreimal an einen anderen Wagen mit einer DAK (Digitalen automatischen Kupplung) gekoppelt. Dabei wurde als Referenz zur Entfernungsmessung ein Maßband zwischen die Schienen gelegt (vgl. Abbildung 5). Im weiteren Projektverlauf wurden im Eisenbahnmuseum Schwarzenberg mehrere Aufnahmen mit und ohne fahrenden Wagon gemacht. Der Unterschied zu den Minden-Aufnahmen ist die fehlende DAK, da dort nur manuelle Scharfenberg-Kupplungen vorhanden sind. Termine in Schwarzenberg waren am:
  - Testfahrten mit Lokführer: 28.08.24; 03.09.24; 10.12.24
  - Testaufnahmen ohne Lokführer: 03.07.24; 22.08.24
  - Besichtigungen: 04.03.24; 02.08.24 (für Workshop)



Abbildung 1: Verwendete Kameras in Schwarzenberg



Abbildung 2: Darstellung mit Personendetektion

- Die zunächst ausgewählten **Kamera-Objektiv**-Kombinationen stellten sich bei den ersten Testaufnahmen in Minden als ungeeignet heraus. Durch die starken Erschütterungen und Vibrationen waren viele Bilder unscharf oder verzerrt und Teile lösten sich. Die Kameras, welche keine PoE-/USB-Industriekameras waren, nehmen nur komprimierte Bilder auf und speichern diese direkt in dem mp4-Videoformat ab. Diese Frames können entweder nicht in Echtzeit unkomprimiert abgerufen werden oder die Komprimierung ruft Schwierigkeiten bei den Bildverarbeitungsalgorithmen hervor. Des Weiteren waren die Helligkeitswert-Anpassungen (Exposure, Gain) während der Datenaufnahmen zu niedrig, sodass während des Kupplungsprozesses die Aufnahmen zu dunkel wurden, weil sich die Lichtverhältnisse zu schnell und zu stark änderten. Für die Aufnahmen in Schwarzenberg wurde eine iDS U3-3270CP-C-HQ Rev.2.2 uEye+ Kamera mit einem KOWA LM3NCM-WP Objektiv (3,5 mm; 1/1.8", Abbildung 1) gewählt. Dieses bietet zwar keine Weitwinkel-Sicht, aber bietet eine permanente Helligkeitsanpassung durch Auto Gain und Auto Exposure. Des Weiteren kann die Region, aus welcher die Pixelwerte für die Helligkeitsanpassung dynamisch (abhängig von Distanz) gewählt werden, sodass durch den hellen Himmel das gesamte Bild nicht unterbelichtet ist. Dadurch kann sich die Belichtungszeiteinstellung auf den wichtigen Bereich der Gleise und Wagons konzentrieren.
- Die Live- bzw. Streaming-Frameübertragung ist mit einer H.264 **Komprimierung** über SPE (praktisch ca. 6 Mb/s) möglich. Des Weiteren wurde der Videostream auf einen

Bildausschnitt zugeschnitten (Abbildung 2), der im Workstream der Digital Rail Convention als Region-of-Interest des Lokführers ermittelt wurde, um die Datenrate auf ca. 2 Mb/s zu reduzieren. Weitere Streaming-Komprimierungsalgorithmen H.265 und AV1 wurden ebenfalls implementiert und haben eine sehr gute Kompressionsrate gezeigt. Diese Verfahren könnten ebenfalls für die Übertragung über SPE verwendet werden. Die Kompressionsrate ist stark vom Bildinhalt abhängig. Bei vielen monotonen Flächen ist die Rate wesentlich größer, da die Quantisierung diese Fläche gut zusammenfassen kann, im Gegensatz zu einer sehr detailreichen Szene, in der die Kompressionsrate sehr gering ausfällt. Das Delay, welches bei der Komprimierung und Dekomprimierung des Frames entsteht, könnte durch eine Hardwareunterstützung verringert werden. Die verwendenden Embedded Systeme der Jeston-Orin-Serie stellen Hardwareunterstützung für H.264 und H.265 bereit und erlauben aufgrund ihrer Leistungsfähigkeit auch weitere Softwarelösungen.



Abbildung 3: SPE-Rack im Labor

- Der **SPE - Laboraufbau** wurde mit 14 Mediaconvertern *SPE zu Ethernet*, verbunden durch Store-and-Forward-Switches und 7 Kabelsegmenten, aufgebaut, siehe Abbildung 3. Die Videoübertragung war bis zu 7 Mb/s möglich. Die Round-Trip-Time (RTT) -Messung wurde im Labor mit diesen 7 Segmenten durchgeführt. Dabei wurde festgestellt, dass die RTT für eine Segmentverbindung (eine Verbindung zwischen 2 Wagons) von der Paketgröße abhängt. Die Erhöhung der RTT durch Anfügen von Segmenten (Switches) hängt von der Framelänge ab, die das System im Store-and-Forward-Prozess verwendet. Bei Paketen, die lange Frames verwenden, erhöht das Hinzufügen eines Segments (Wagens) die RTT um ca. 2-3 ms (abhängig von der Länge des Frames), was insgesamt 15 ms für die Nutzung von 6 Segmenten ergibt. Für den 7-Segment-Aufbau wurde eine RTT von 18 ms bei einer Paketlänge von 1518 Bytes beobachtet. Bei kurzen Paketen (weniger als 1518 Bytes) betrug die RTT-Erhöhung mit der Nutzung von 6 Segmenten (6 Switches) 3 ms. Dies ist ein kleiner Wert, da die Frame-länge kleiner als bei längeren Paketen ist.
- Zur weiteren Verringerung der Datenrate können sowohl die **Framerate** als auch die **Auflösung** verringert werden. Subjektiv hat sich ergeben, dass eine Auflösung von 540x480px oder 720x640px ausreichen, um die wichtigen Merkmale im Bild, wie Gleise, Kupplung und Wagons, zu erkennen. Die Framerate sollte optimal bei 20 fps (frames per second) liegen, um kein bemerkbares Ruckeln auf dem Display des Lokführers zu erzeugen. Da auch ein Frame mit einer größeren Auflösung übertragen werden kann, wird das Bild in Originalgröße zuzüglich Metadaten wie Distanz, Geschwindigkeit und Bounding Boxen von Personen übertragen.
- Damit die entwickelten Algorithmen funktionieren, muss zuerst die Gleiserkennung erfolgreich ausgeführt werden. Mit den detektierten Gleisen kann die Region-of-Interest (die Region der Schwellen für die Geschwindigkeitsschätzung) festgelegt werden.

Durch den genormten Gleisabstand von 1435 mm kann man ermitteln, wieviel Millimeter einem Pixel entsprechen. Somit hat man immer einen Bezug zur realen Welt. Für die Gleiserkennung wurden zwei verschiedene Herangehensweisen betrachtet: Erstens die Hough Line Transformation und zweitens die Bezier Curve Extraction. Die Hough Line Transformation hat den Vorteil, dass sie wesentlich zuverlässiger als die Bezier Funktion, speziell bei geradem Gleisverlauf, ist. Dennoch gab es einige falsch-positive Detektionen des Gleises verursacht durch das Maßband im Gleisbett, welches bei unseren Test-Aufnahmen für die realen Distanz- und Geschwindigkeitswerte benötigt wurde. Auch ohne Maßband im Gleisbett gab es fehlerhafte Detektionen durch den Schatten der Schienen, welches aber einen geringeren Einfluss auf die Distanzdetektion hat (Abbildung 4). Durch die neue feste Position und Winkel der Kamera können Intervalle der x-Koordinate und Anstieg des linken und rechten Gleises verwendet werden, um deren Bestimmung zu verbessern. Die Bezier-Gleiserkennung ist, aufgrund des ähnlichen Grauwertes des Schotters im Gleisbett und der Gleise selbst (Abbildung 5), instabil. Vor allem im Nahbereich des anzukuppelnden Wagons ist diese sehr ungenau, was sich negativ auf die Distanzschätzung auswirkt. Des Weiteren benötigt diese Funktion die x-Koordinate, wo sich die Schienen im unteren Bildrand befinden, im Gegensatz zu der Hough Line Transformation, die die Gleise ohne diese x-Koordinate bestimmen kann.



Abbildung 4: Gleiserkennung der Hough-Line-Transformation

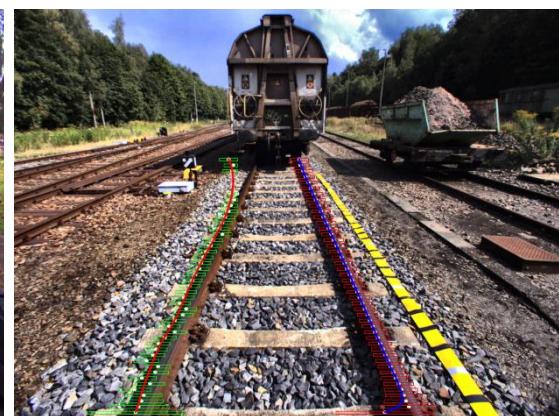


Abbildung 5: Gleiserkennung mit Bezier Curve Extraction

- Die **Geschwindigkeitserkennung** der Lok funktioniert zurzeit über eine Schwellenerkennung und -verfolgung. Durch den realen Wert des Gleisabstandes von 1435 mm kann ermittelt werden, wie viele Millimeter ein Pixel in diesem Szenario sind. Mit diesem mm/px-Wert und die Anzahl der Pixelverschiebung einer Schwelle in einem Zeitraum von einer Sekunde (bestimbar mit fps der Kamera) kann die reale Geschwindigkeit linear bestimmt werden. Dieser Wert kann durch die Verwendung von mehreren Schwellen realitätsnaher durch den Mittelwert ermittelt werden. Die Schwellenerkennung wurde mittels Hough Line Transformation und Line Segment Detector (LSD) durchgeführt. Eine weitere Möglichkeit zur Geschwindigkeitsschätzung ist der Optical Flow, welche die Verschiebung eines Pixels von Frame 1 zu Frame 20 nach einer Sekunde (20 fps) bestimmen kann. Unter Verwendung mehrerer Pixel kann hier ebenfalls ein Mittelwert bestimmt werden, welcher zur Geschwindigkeitsschätzung verwendet wird. Diese Methode bezieht sich auf einen vordefinierten Bildausschnitt des Gleises und Wagons, um eine genauere Geschwindigkeit berechnen zu können. Alle Methoden haben eine bessere Performanz bei größeren Distanzen, aber sind sehr ungenau während des Kupplungsprozesses (geringe Distanz), da die Schwellen schwer

detektierbar oder nicht mehr sichtbar sind. Der Mean Average Error (MAE) unter Verwendung der Hough Schwellendetektion betragen 0,78 km/h (iDS); 0,49 km/h (ELP) und 0,77 km/h (HIK). Je nach Objektiv- und Kamera-Einstellungen funktionieren die Berechnung unterschiedlich gut. Das ELP-Video (Aufnahme 1) hat mit unterschiedlichen Methoden folgende MAE: 0,49 km/h (Hough), 1,13 km/h (LSD) und 0,56 km/h (Optical Flow). Unter Betrachtung der Diagramme und dieser Ergebnisse liefert der Optical Flow die besten Ergebnisse bis zu einer bestimmten Distanz zu dem anzukuppelnden Wagen. Im Nahbereich ist auch diese Methode sehr ungenau und fehleranfällig. Dennoch sind alle Ergebnisse der Geschwindigkeitsschätzung sehr sprunghaft, da es zu Fehlern bei der Schwellenzuordnung in den darauffolgenden Frames kommt, wenn eine Schwelle im unteren Bildrand verschwindet. Die Schwellenerkennung besitzt selbst einige Fehlinterpretationen, welche durch das Maßband im Gleisbett hervorgerufen wird. Im Laufe des Projektes sollen diese Fehler behoben oder eine neue Geschwindigkeitsschätzung implementiert werden.

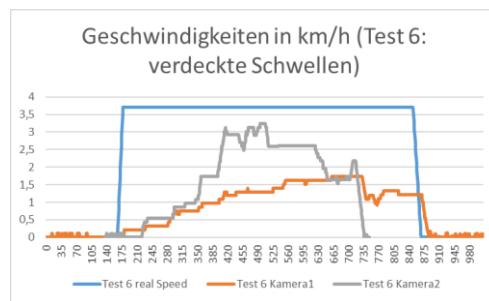


Abbildung 6: Geschwindigkeit in km/h zu Frames für Test 6, indem einige Schwellen zu Beginn verdeckt durch Erde waren

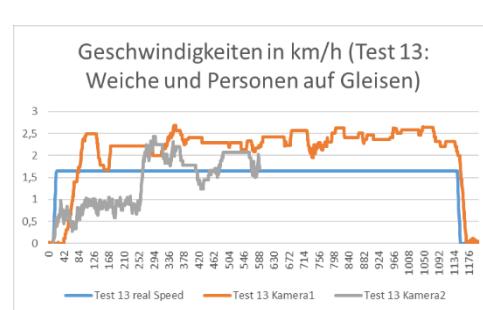


Abbildung 7: Geschwindigkeit in km/h zu Frames für beide Kamerapositionen mit Weiche und Personen auf den Gleisen

- Mit einer weiteren angepassten Geschwindigkeitsschätzung, welche ebenfalls LSD zur Schwellendetektion verwendet und die Anzahl der Schwellen verringert auf 3 (Schwellen), kann eine Genauigkeit von MAE 0,475km/h bei hellem Schotter erreicht werden (Kameraposition 1: 50° und Höhe 71cm). Eine höhere Kameraposition (103cm) mit flacherem Winkel von 40° dagegen 0,648 km/h MAE. Dennoch spielt bei diesem Verfahren die Schwellensichtbarkeit eine große Rolle, da bei verdeckten Schwellen der MAE bei beiden Kameras über 2km/h beträgt (Abbildung 7). Bei Weichen sind die Schwellen unterbrochen und haben Fehler in der durchgängigen Detektion. So kommt die MAE auf 0,862 km/h (Kamera1) und 0,953 km/h (Kamera2). Personen im oder neben dem Gleis haben keinen Einfluss auf die Geschwindigkeitsschätzung, welche auch funktioniert, wenn kein Wagon sichtbar ist (Abbildung 7).
- Die **Distanz zwischen** dem fahrenden und dem stehenden anzukuppelnden Wagon kann mittels Bildverarbeitungsmethoden bestimmt werden. Als fahrender Wagon, an dem die Kamera angebracht ist, wird der untere Bildrand angesehen und der anzukuppelnde Wagon ist in dem Frame zu sehen. Zuerst muss die Position des zu kuppelnden Wagens bestimmt werden, welches mit einer einfachen Wagonerkennung mit Template Matching und Tracking durchgeführt wurde. So konnte die y-Koordinate bestimmt werden, an welchem das Gleis endet und der Wagon beginnt. Diese y-Koordinate ist der Ausgangspunkt für die Distanzschätzung, da der Gleisabstand an dieser Stelle oder die Höhe dieser Stelle für die Distanzschätzung verwendet wird. Das Template Matching hat jedoch den Nachteil, dass bei einem anderen Wagon die Distanzen nicht mehr berechnet werden können, da diese y-Koordinate fehlt. Für ein darauffolgen-

des Projekt könnte eine vollwertige Wagonerkennung mittels eines CNNs implementiert werden, wenn genügend gelabelte Bild-Daten vorhanden sind oder aufgenommen werden können.

Anschließend kann die Distanz zu dem zu kuppelnden Wagen über 4 unterschiedliche Berechnungen gewonnen werden (Tabelle 1).

Distanz-berechnung	Gleichung	Bemerkung
Focal Length	$y = \frac{1435\text{mm} * \text{focalLength}}{\text{pxAnz}} * 0,001$	Angepasst von State of the art
Linear Width Railway	$y = 0,887 * \left(\frac{1435}{\text{pxAnz}}\right) - 2,7$	y ist Distanz in m und pxAnz der Breite des Gleises an dem anzukuppelnden Wagon
Linear Height Railway	$y = 0,62 * (x) - 2,7$	y ist Distanz in m und x ist mm/px
	$y = 0,62 * \left(\frac{y * 1000}{p}\right) - 2,7$	*1000 wegen Umrechnung von mm in m
	$y = -2,7 * \left(\frac{p}{p - 620}\right)$	
Exponential Height Railway	$y = 0,062 * e^{0,0111 * \text{pxHeight}} + 0,608$	

Tabelle 1: Möglichkeiten zur Distanzberechnung

Die erste Möglichkeit zur Distanzschatzung ist die Brennweitengleichung, welche die bekannte Brennweite des Objektivs, die Pixelanzahl zwischen den Gleisen bei dem zu kuppelnden Wagen und den realen Gleisabstand von 1435 mm verwendet. Die zweite Möglichkeit ist der ermittelte lineare Zusammenhang zwischen Gleisabstand und Pixelanzahl. Je geringer die Distanz, desto größer ist die Pixelanzahl zwischen den Gleisen. Die lineare Gleichung ist je nach verwendeter Kamera und Objektiv unterschiedlich und diese Gleichungen mussten angepasst bzw. neu bestimmt werden. Die letzten beiden Gleichungen benutzen den linearen und exponentiellen Zusammenhang zwischen dem Abstand und der Pixelanzahl ( $p = \text{pxHeight}$ ) der gesuchten Distanz, welches sich aus den y-Wert an dem Wagon und an dem unteren Bildrand zusammensetzt. Diese haben je nachdem, ob die Pixelanzahl in px oder die Millimeter pro Pixel als Eingabewert der Gleichung verwendet wurde einen anderen Zusammenhang. Der exponentielle Zusammenhang bei mm/px musste anschließend nach der Distanz umgestellt werden. Auch hier ändern sich die mathematischen Gleichungen, je nachdem, welche Kamera-Objektiv-Kombination verwendet wurde.

Je nach Kamera und Objektiv ist eine andere Berechnungsart genauer. So verwendet ELP die lineare Breitengleichung und hat einen MAE von 0,82 m. Die HIKVISION Kamera hat einen MAE von 0,56 m und benutzt die lineare Längengleichung des Gleises und iDS hat zwar eine sehr hohen MAE von 1,92 m (lineare Breitengleichung des Gleises), ist dafür bei einer Distanz unter 5m am genauesten. Unter einem Abstand von 10 m wurde die Distanzschatzung insgesamt bei allen Kameras hinreichend genau, im Gegensatz zu der Anfangsdistanz von 23-28 m. Dies liegt an der größeren Pixelanzahl zwischen den Gleisen in der Nähe des anzukoppelnden Wagons (entspricht der Spurbreite von 1435 mm). Ein Pixel zwischen den Schienen hat bei einer längeren Distanz eine größere Auswirkung auf die Distanzberechnung, weil dieses Pixel in der Realität

einer größeren Distanz entspricht, im Gegensatz zu einer geringen Distanz, bei welcher ein Pixel weniger Millimetern entspricht.

Das Maß des kreisförmigen Greifbereiches des Kupplungskopfes der Scharfenberg-Kupplung ist standardisiert und hat einen Durchmesser von 210 mm. Unter Verwendung der Kreiserkennung Hough Circle Transformierung und der Brennweitengleichung konnte mit diesem Wert die Distanz im Nahbereich unter 5 m bestimmt werden. Diese Distanzschätzung brachte keine höhere Genauigkeit, als die Bestimmung mit dem Gleisabstand. Des Weiteren benötigen diese Kreisbestimmung und Berechnung sowohl mehr Rechenleistung und –zeit, welche zu einem größeren Delay im Nahbereich führt, als dies bei der vorherigen Distanzschätzung der Fall ist. Aus diesem Grund wurde diese Distanzschätzung in das System nicht mit implementiert.

Die oben beschriebenen Tests fanden an einer DAK statt, indem eine Genauigkeit von 0,5m im Nahbereich ausreichend ist. Um die Anwendung auf weitere Kupplungsarten zu übertragen, sollte die Genauigkeit höher sein, da sonst die Kupplung selbst beschädigt werden kann, falls mit zu hoher Geschwindigkeit gekuppelt wird. In den nachfolgenden Tests wird die Scharfenberg-Kupplung als Testkupplung verwendet.

Als Wagondetektion wurde das YOLOv8s Modell verwendet, um gleichzeitig Personen zu detektieren und die Rechenzeit so gering wie möglich zu halten. Dieses Modell schwankt etwas bei der Detektion der unteren Kante des Wagons, welche für die Distanzschätzung verwendet wird. Dieses Problem kann aber durch eine Mittelwertbestimmung der letzten 8 Distanzen kompensiert werden. Da es im Nahbereich (ab 3m) beim YOLOv8s-Modell (s-small) Schwierigkeiten bei der Detektion des Wagons gibt und dieser nicht mehr detektiert wird, musste auf ein größeres YOLOv8l-Modell (l-large) aufgestockt werden. Dieses größere Modell kann den Wagon auch im Nahbereich detektieren und führt zu einer genauen Distanz solange sich die Gleise im Frame befinden. Bei korrekt erkannten Gleisen unter Laborbedingungen ist der gesamte MAE 0,99m bei einer maximalen Distanz von 19m und im Nahbereich von 7m 0,26m (MAE) und im weiteren Nahbereich bis 3m ist MAE 0,22m (Abbildung 8). Falls es zu falsch detektierten Gleisen kommt, sind die MAE 1,3m (bis 19m); 0,5m (bis 7m) und 0,28m (bis 3m).

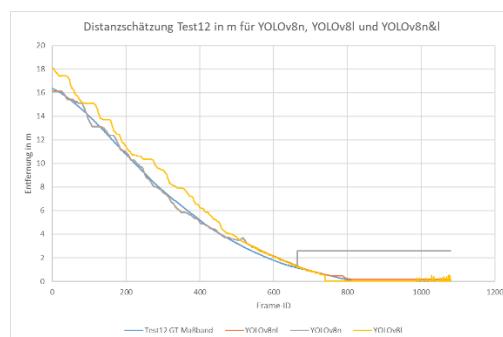


Abbildung 8: Distanzschätzung in m für Test ohne Helligkeitsabpassung im Nahbereich für Methode mit 3 grünen Rechtecke

Laut Bahnangestellten ist im Nahbereich ein Fehler von  $\pm 0,28m$  zu groß. Da bei einigen dieser Nahbereich-Frames der Wagon nicht mehr komplett im Frame sichtbar ist und somit die Verbindung zur Real-Welt, das Gleis, aus dem Sichtfeld der Kamera ver-

schwunden ist, muss die Distanz anders bestimmt werden. Es wurden 3 grüne rechteckige Magnete mit bekannten Abständen zueinander (39cm; 18,5cm und 20cm) an den anzufahrenden Wagon befestigt. Anschließend wurden die Eckpunkte der Rechtecke durch einen Farb-Schwellwert/Chroma-Keying detektiert. Anhand der Brennweite des Objektivs, der Pixelanzahl und realen Abständen kann die Distanz mit der Brennwert-Gleichung berechnet werden (Abbildung 9). Da es sich hier um drei Rechtecke handelt, können 3 verschiedene Distanzen bestimmt werden und der Mittelwert dieser ist der neu kalkulierte Abstand der beiden Wagons. Mithilfe dieses Verfahrens wird im Nahbereich ab 3m ein mAP von 0,14m erzielt werden (Abbildung 10).

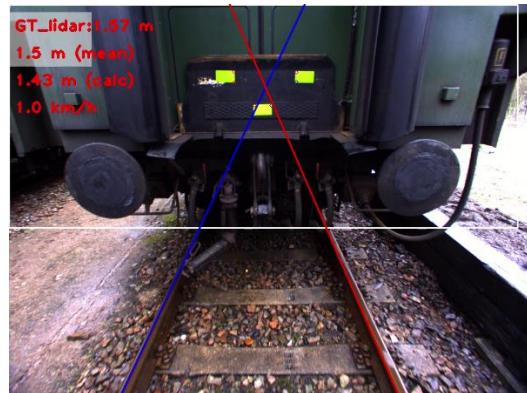


Abbildung 9: Sicht der Kamera mit eingezeichneten Ergebnissen der Algorithmen (Gleiserkennung, Wagondetektion, Eckpunktermittlung der Rechtecke) mit GroundTruth-Distanz (GT-lidar), gemittelter Distanzwert (mean), berechneter Distanz im jetzigen Frame (calc) und Geschwindigkeit des angezeigten Frames

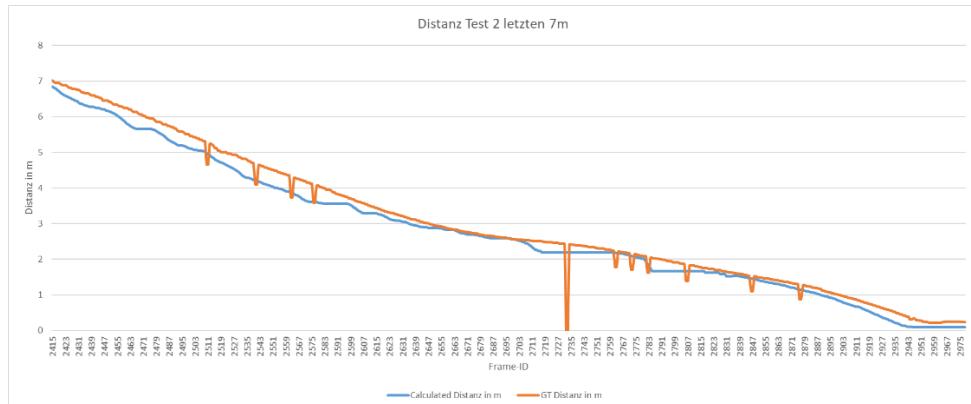


Abbildung 10: Distanz im Nahbereich (letzten 7m vor Kupplungsprozess) mit Helligkeitsanpassung

- Die **Personendetektion** im Gefahrenbereich geschah mittels Faster RCNN, welches mit den aufgenommenen Daten aus Minden eine mAP (mean Average Precision) von 0,91 aufweist. YOLOv5 lieferte genauere Ergebnisse mit einer mAP von 0,994. Zusätzlich zu der detektierten Person wird dessen Bounding Box Abstand zu den Gleisen berechnet, indem die Pixelanzahl der unteren Ecke der Bounding Box zu den Gleisen, der Pixelwert des Gleisabstandes und der reale Gleisabstand von 1435 mm verwendet

wird. Wenn der Abstand geringer als 1 m ist, wird eine Gefahrenmeldung in dem Frame mit angezeigt.

Für die Personendetektion wurde ebenfalls ein YOLOv8s Modell verwendet, welches ebenfalls echtzeitfähig ist und eine ähnliche Genauigkeit aufweist.

- Die **Region-of-Interest (RoI)** besagt, welcher Bildausschnitt für den Lokführer in diesem Moment interessant ist und übertragen werden soll, damit nicht das gesamte Frame übertragen werden muss, sondern lediglich ein Ausschnitt. Der Grund für das Ausschneiden des Bildes ist die Verringerung der Bilddatenrate. Je nach geschätzter Distanz ändert sich die Region-of-Interest. Das Ergebnis des Workstreams ergab, dass bei einer Distanz größer als 10m das Gleis und nicht der Wagon von Interesse ist und bei einer kleineren Distanz kann der Wagon mit der Kupplung und Gleis übertragen werden. Wenn sich eine Person auf den Gleisen oder in deren Nähe befindet, ändert sich dieser übertragene Bildausschnitt ebenfalls, indem die Person in Relation zu dem Gleis als Frame übermittelt wird.
- Die oben genannten Bildverarbeitungsalgorithmen, die Komprimierung und die Datenübertragung benötigen Laufzeit bis das Frame bei dem Lokführer ankommt. Diese Laufzeit wird als Delay bezeichnet. Um dies auszugleichen, wird eine **Delaykompensation** mittels eines Zooms durchgeführt. Wenn die Dauer des Delays und die Geschwindigkeit berechnet wurde, kann bestimmt werden, welche Distanz der Lokführer gefahren ist, bis das Frame in dem Human-Machine-Interface im Führerstand angezeigt wird. Diese Distanz wird in Pixel umgerechnet und diese Pixelanzahl wird an dem unteren Bildrand der Region-of-Interest ausgeschnitten. Dadurch hat der Lokführer in seiner Fahrerkabine eine Simulation des aktuellen Frames der Rückfahrkamera, welche sich am Ende des Zuges befindet.
- Die Implementierung der oben genannten Algorithmen erfolgte auf **Embedded Systeme**. Hierfür wurden Embedded Systeme vom Typ Nvidia Jetson Orin AGX Developer Kit ausgewählt, da die Jetson-Orin-Serie die nötige Rechenleistung insbesondere für Bildverarbeitungsalgorithmen zur Verfügung stellt und dabei klein und stromsparend ist. Durch die Verwendung der Developer-Kit-Version stand ein voll funktionsfähiges Embedded System zur Verfügung, so dass keine zeit- und kostenaufwendige Entwicklung spezieller Hardware, wie z.B. Leiterplatten, mehr erforderlich war. Zusammen mit einer Kamera stellte der Jetson Orin AX einen Smart Sensor dar, welche Live-Bilder aufnahm, verarbeitete und die verbreiteten Daten in Form von Metadaten (Geschwindigkeit, Distanz, detektierte Personen) sowie das zugeschnittene Bild für die Rückfahrkamera über SPE versendete.
- Die nötigen Optimierungen für die Embedded Systeme wurden hauptsächlich von der SYS TEC durchgeführt, so dass im September ein erster Demonstrator bereitstand. Im Rahmen umfangreicher Tests wurden noch einige Verbesserungsmöglichkeiten ermittelt und in Form der finalen Algorithmen implementiert. Im Projektverlauf gab es im Zeitplan immer wieder Verzögerungen, so dass die finalen Algorithmen nicht mehr auf den Smart Sensor übertragen und damit in Form eines Demonstrators vorgeführt werden konnten. Einer beantragten, kostenneutralen Projektverlängerung um wenige Monate zur Fertigstellung des finalen Demonstrators wurde leider nicht stattgegeben.
- Als **HMI** (Human-Machine-Interface), also der Benutzerschnittstelle für den Lokführer, wurde ein Display und ein Embedded System von Typ Nvidia Jetson Orin Nano verwendet. Diese kleinere, leistungsschwächere Version des Orin AGX reicht aus, um die Daten via SPE zu empfangen, auszuwerten und auf dem Display anzuzeigen. Dieses (Software-) HMI wurde für Laboraufbauten, Tests und später als zusätzliches Display im Demonstrator verwendet.
- Neben diesem selbstentwickelten (Software-) HMI wurde während des Projektes ein weiteres (Hardware-) HMI zur Verfügung gestellt. Dabei handelt es sich um ein fertiges, bahn zugelassenes Produkt (HMI 84ab-TW/SIL von Selectron Systems AG), welches

heute schon in vielen Zügen Verwendung findet. Dieses wurde dann im Demonstrator verwendet, was die Praxistauglichkeit des Systems weiter unterstrich. Da die damals verfügbaren HMIs mit Bahnzulassung noch keine modernen Videokompressionsverfahren unterstützten, musste auf den bewährten H.264-Standard für das Live-Video zurückgegriffen werden. Für die Embedded Systeme stellte dies kein Problem und auch die SPE-Datenverbindung konnte die dadurch geringfügig gestiegene Datenrate problemlos handhaben.

- Für die Bewertung des 5G-bezogenen Leistungspotenzials für die Videoübertragung in Echtzeit wurden zwei 5G-Evaluierungsboards des Modells M2.5G erworben. Jede Platine war mit einem 5G-Modul ausgestattet. Die Platinen mussten leicht modifiziert werden, um die Module und die Kühlkörper unterzubringen. Geeignete Antennen für das 5G-Band 3,5 GHz wurden an den Boards montiert und 5G-SIM-Karten von Vodafone verwendet. Anschließend wurden die Treiber für die Boards installiert und beide Boards mit AT-Befehlen konfiguriert. Die Konfiguration umfasste das Testen verschiedener Modi (LTE, 5G, 5G Standalone) sowie die Konfiguration der Verbindung zwischen den Boards und den Host-PCs über USB. Beide müssen sich im selben Netzwerk befinden. Daher wurde der Datenverkehr mit einer VPN-Verbindung für jeden Host-PC getunnelt.

Die Verbindung zwischen den beiden Boards wurde mit zwei Methoden getestet: Eine quantitative und eine qualitative. Für die quantitative Methode wurde JPerf, eine GUI für das Open-Source-Geschwindigkeitstest-Tool iPerf, verwendet. Ausgewertet wurden die Datenrate (für TCP- und UDP-Pakete), der Delay-Jitter sowie die Drop-Rate (für UDP-Pakete). Die TCP-Tests geben Aufschluss über die erreichbare Datenrate für den Datenverkehr im Allgemeinen und wurden daher zum Vergleich von LTE (4G) und 5G herangezogen. Die UDP-Tests sind besser geeignet, um sich ein Bild vom Verhalten von Datenklassen zu machen, für die es keine Kontrollmechanismen gibt, wie z.B. bei Video-Streaming-Anwendungen. Der TCP-Test für LTE ergab eine Datenrate zwischen 3 Mbit/s und 72 Mbit/s, wobei der Durchschnitt bei 30 Mbit/s lag (Abbildung 11). Die 5G-Datenrate hingegen war während des Testzeitraums viel stabiler mit einem Durchschnitt von 43 Mbit/s und viel geringeren Schwankungen (mindestens 18 Mbit/s und höchstens 68 Mbit/s), wie in Abbildung 12 zu sehen ist.

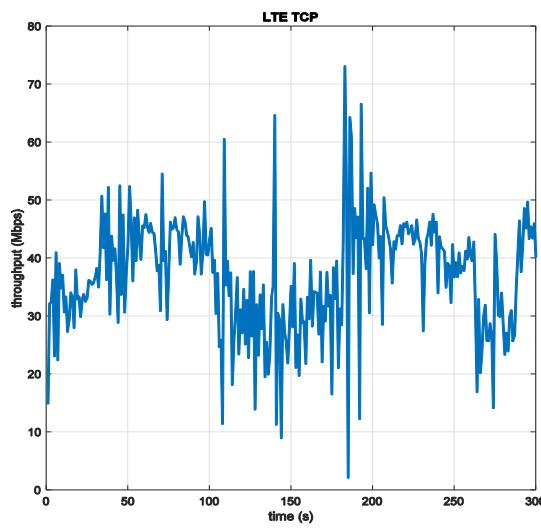


Abbildung 11: TCP Datenrate LTE

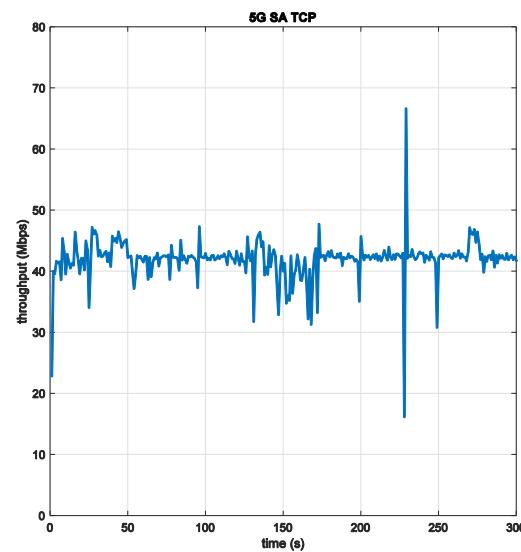


Abbildung 12: TCP Datenrate 5G

Da UDP-Pakete nicht über einen Rückkopplungsmechanismus verfügen, ist der Test schwieriger umzusetzen als mit TCP. Die Datenrate muss auf der Serverseite (PC1)

manuell gewählt werden und der Client (PC2) meldet, was er empfangen hat. Zu Beginn wurde eine hohe Datenrate von 50 Mbit/s gewählt und dann schrittweise um jeweils 10 Mbit/s verringert. Wie in Abbildung 13 dargestellt, wurde bei 50 Mbit/s eine hohe Abbruchrate beobachtet. Die Abfallrate blieb bis Erreichen von 10 Mbit/s hoch, wo sie minimal war und der entsprechende Jitter unter 2 ms lag, wie in Abbildung 14 dargestellt.

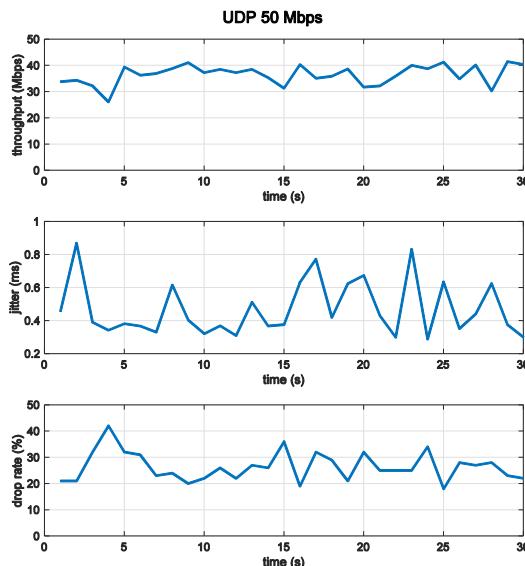


Abbildung 13: UDP Parameter 50 Mbit/s

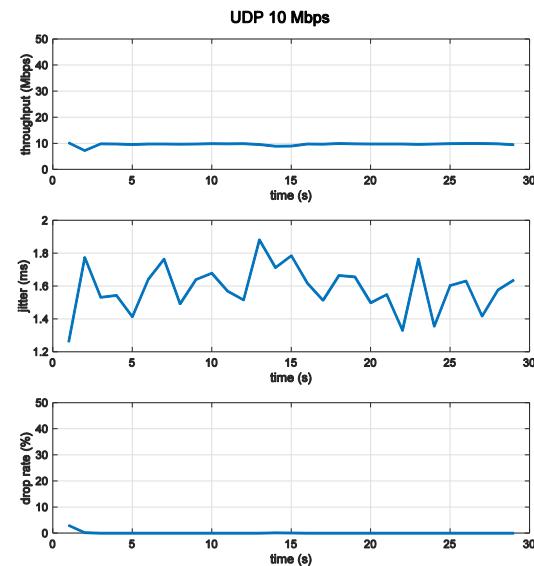


Abbildung 14: UDP Parameter 10 Mbit/s

Bei der qualitativen Methode wurden mehrere Videos, die während des Projekts mit unterschiedlichen Kameras, Auflösungen, Bitraten und Frameraten aufgenommen wurden, mit dem VLC-Player vom Server zum Client gestreamt. Die Streaming-Qualität wurde manuell beurteilt und jede Unterbrechung als Streaming-Fehler gewertet. Tabelle 2 zeigt eine Zusammenfassung der Ergebnisse.

Video Titel	Bitrate	Erfolg
iDS1-1920x1080px-20fps	453 kbps	J
HIKVISIONAufnahme1-1620x1440px-20fps	6095 kbps	N
Hikvision-5120x1440-20fps-5Mbits	5142 kbps	N
ELPaufnahme2-1280x720px-20fps	1145 kbps	J
Dashcam4-3840x2160px-20fps	6065 kbps	N

Tabelle 2: Videostreaming mit 5G

Die Ergebnisse von 5G-Verbindungstests über das öffentliche Netz auf dem Universitätsgelände zeigen, dass trotz der Verbesserungen der 5G-Datenraten im Vergleich zu 4G die Anforderungen für hochauflösendes Videostreaming noch nicht erfüllt werden. Dennoch kann 5G eine praktikable Option als Backup für drahtgebundene Verbindungen mit reduzierter Auflösung und Framerate darstellen. Für die zukünftige Anwendung im Bahnbereich kann mit der Einführung von FRMCS eine höhere zuverlässige Übertragung erwartet werden.

## 2 Wichtigste Positionen des zahlenmäßigen Nachweises

Das Projekt wurde an der TU Chemnitz unter der Leitung des ZWT durchgeführt. Die technischen Arbeiten der Arbeitspakete seitens der TU fanden an der Professur Digital- und Schaltungstechnik statt. Die im Rahmen des Projektes entstandenen Ausgaben betragen insgesamt 570.887,36 € (zzgl. Projektpauschale i. H. v. 114.177,47 €). Sie liegen damit 49.592,52 € (ohne Projektpauschale) unter der bewilligten Zuwendung. Die Ausgaben wurden im Wesentlichen für Personalausgaben zur Beschäftigung wissenschaftlicher Mitarbeiter (301.406,10 €), für die Ausgaben zur Vergabe von Aufträgen (233.558,00 €) sowie für Gegenstände über 800 € (29.248,09 €) verwendet. Weiterhin erfolgten Ausgaben für Gegenstände bis 800 € (1.312,19 €), für Mieten und Rechnerkosten (Tests/Testfahrten; 4.872,78 €) sowie für Dienstreisen (490,20 €).

Insbesondere die Positionen Mieten und Rechnerkosten (0834) und Dienstreisen (0846) fielen geringer aus als ursprünglich geplant.

Die Systemtests konnten erst in den letzten Projektmonaten durchgeführt werden. Im Zuge dessen wurden noch Verbesserungsmöglichkeiten ermittelt. Die finalen Algorithmen konnten während der Projektlaufzeit nicht mehr auf den Smart Sensor übertragen und damit in Form eines Demonstrators vorgeführt werden. Aus diesem Grund konnten die Mittel für die Anmietung eines Forschungszuges nicht im ursprünglich geplanten Umfang innerhalb der Projektlaufzeit verausgabt werden.

Die Reisekosten wurden nicht vollständig ausgeschöpft. Dienstreisen wurden nach Minden und Annaberg-Buchholz sowie nach Frankfurt zu einer Expertenrunde im Rahmen der Initiative Shift2Rail gemacht. Zudem wurden Projektbesprechungen als Online-Abstimmungen durchgeführt, was zu einem geringeren Bedarf an Dienstreise-Mitteln führte.

## 3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Zu Projektbeginn wurde mit an der Universität vorhandenen Kameras die Umsetzbarkeit ausprobiert, um Kosten zu sparen. Nachdem sich herausstellte, dass diese ungeeignet sind, wurden weitere Kameras mit anpassbaren Parametern bestellt und verwendet, wie es in der Ressourcenplanung vorgesehen war. Ebenso wurde mit den Embedded Systemen, Display und der Workstation verfahren, welche für die Algorithmenentwicklung verwendet wurden. Sofern sich herausstellte, dass neue Hardware notwendig war, wurde diese im Rahmen des geplanten Budgets bestellt und verwendet.

Von den geplanten Projektarbeiten wurden zunächst alle umgesetzt. Der dafür vorgesehene Aufwand war angemessen. Im späteren Projektverlauf wurde die Latenzkompensation nicht weiter verwendet.

## 4 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die Ergebnisse finden Anwendung in der universitären Lehre.

Die Projektergebnisse bilden die Grundlage für die Entwicklung und Akquise von Folgeprojekten bzw. von weiteren Forschungsvorhaben, die inhaltlich auf dem vorliegenden Projekt aufbauen.

Des Weiteren ist geplant, die Projektergebnisse auf wissenschaftlichen Konferenzen und Messen, wie der Digital Rail Convention im September 2025, vorzustellen.

## 5 Erfolgte oder geplante Veröffentlichungen

Im Rahmen der Digital Rail Convention im September 2023 wurde durch die Professur Digital- und Schaltungstechnik ein dreistündiger Workstream mit dem Titel „Rückfahrkameras an Güterwagen - Videoübertragung durch bestehende Bussystem eines Güterwagens“ angeboten. Der Workstream umfasste neben einer Vorstellung des Projektes auch die Möglichkeit zur Diskussion mit den teilnehmenden Personen. Insbesondere sollte mit den Teilnehmenden erörtert werden,

wie eine Rückfahrkamera entsprechend den Anforderungen von Zugführern und Bahnangestellten optimiert werden könnte und mögliche Schwachstellen identifiziert werden.

Des Weiteren fand am 08.02.2024 die feierliche Eröffnung der 5G-Forschungsinfrastruktur durch die Technische Universität Chemnitz und die Deutsche Bahn AG am Forschungscampus in Annaberg-Buchholz statt. An dieser Veranstaltung nahmen Gäste aus den Bereichen Politik, Bildung und Wirtschaft teil, wie z. B. Dr. Volker Wissing (Bundesminister für Digitales und Verkehr), Michael Kretschmer (sächsischer Ministerpräsident), Prof. Dr. Gerd Strohmeier (Rektor der TU Chemnitz) und Dr. Daniela Gerd tom Markotten (Vorständin für Digitalisierung und Technik der Deutschen Bahn AG). Neben anderen WIR!-Projekten konnte während eines Rundgangs auch das Projekt „ReViCaT“ vorgestellt werden.

Zudem ist eine Vorstellung der Projektergebnisse in Form einer Demonstration im Rahmen der Digital Rail Convention im September 2025 in Planung.

Auf der Webseite des Smart Rail Connectivity Campus sind ebenfalls Informationen zum Projekt veröffentlicht (Link: <https://www.smart-rail-campus.de/?p=6288>).



Abbildung 15: Projektvorstellung durch Herrn Prof. Dr. Hirtz und Herrn Sieber (Sys Tec) während des Rundgangs des Verkehrsministers im Zuge der Eröffnung der 5G-Forschungsinfrastruktur