

Sachbericht zum Verwendungsnachweis

Teil I: Kurzbericht

Verbundprojekt

Multimodale Erfassung des Normalenvektorfelds Transparenter Objekte für Roboter
MENTOR

Teilprojekt

Multimodale Bilderkennung von transparenten und semitransparenten Objekten für den industriellen Einsatz

Förderkennzeichen

13N16522

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Laufzeit

01.12.2022 - 31.03.2024

Zuwendungsempfänger

Basler AG
An der Strusbek 60-62
22926 Ahrensburg



Projektleiter

Herr Dr. Ralf Lay (bis 11/2023)
Herr Dr. Jörg Kunze (ab 12/2023)

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des [Bundesministeriums für Bildung und Forschung](#) unter dem Förderkennzeichen 13N16522 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

1 Ursprüngliche Aufgabenstellung / wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Ziel des Projekts war es, die Erkennungsrate und die Genauigkeit der Posenbestimmung von transparenten Objekten wie Glas signifikant zu verbessern. Dies sollte durch die Entwicklung eines multimodalen Bildverarbeitungssystems erreicht werden, das verschiedene Sensortypen und fortschrittliche Algorithmen kombiniert. Ein zentrales Anliegen war die Integration und Synchronisation von Sensordaten aus unterschiedlichen Quellen, um eine zuverlässige und schnelle Objekterkennung zu ermöglichen.

Es gab zwar bereits vor Projektbeginn einige Ansätze zur Detektion und Segmentierung transparenter Objekte, jedoch beschränkten sich diese meist auf die Nutzung von Farbsensoren und Tiefensensoren. Andere Modalitäten wurden kaum kombiniert, und es existierten keine Lösungen für eine multimodale Sensordatenfusion, die auch unter variablen Industriebedingungen zuverlässig arbeiten.

In der industriellen Praxis existierten zudem keine Algorithmen, die eine für industrielle Anwendungen geeignete Fusion und Verarbeitung der Informationen zur Erkennung transparenter Objekte mit verschiedenen optischen Modalitäten ermöglichten. Auch waren die bekannten Systeme oft komplex und in vielen Bereichen nur eingeschränkt einsetzbar. Eine umfassende und zuverlässige Lösung zur Erkennung und Handhabung transparenter Objekte stand nicht zur Verfügung.

2 Ablauf des Vorhabens

Es war geplant, ein multimodales Bildverarbeitungssystem zur Erkennung und Handhabung transparenter Objekte zu entwickeln. Dazu sollten in einer Definitionsphase Kenngrößen und Ziele für die Systemleistung festgelegt und die Anforderungen an die Laborautomation sowie weitere Anwendungen analysiert werden. Zeitgleich waren theoretische Vorarbeiten zur Schnittstellenentwicklung geplant, um ein reibungsloses Zusammenspiel der Systemkomponenten sicherzustellen. Anschließend sollte der technische Sensorkopf ausgelegt werden. Dabei waren auch Auswahl und Integration der verschiedenen Sensoren geplant, wobei thermische, mechanische und elektrische Parameter berücksichtigt werden mussten. Die Analyse und Auswahl der Imaging-Sensoren sowie die Vorbereitung der Algorithmus-Entwicklung bildeten dabei einen zentralen Schwerpunkt. Anschließend sollten Algorithmen zur Erkennung transparenter Objekte durch Kombination klassischer und KI-basierter Verfahren zur Verarbeitung multimodaler Signale realisiert werden. Ziel der Arbeiten war ein Konzept zur Erzeugung hybrider Lösungen und die Realisierung eines Software-Minimaldemonstrators auf Basis der Basler-Software Pylon. Damit sollte die Erprobung der theoretischen Ergebnisse in realen Anwendungen durchgeführt werden.

Von dieser Planung mussten wir aus verschiedenen Gründen deutlich abweichen.

Nach Anmeldung der Insolvenz des Projektpartners Spheron übernahmen wir in Absprache mit den Projektpartnern Arbeitspakete von Spheron, die von zentraler Bedeutung für das Gesamtvorhaben waren. Dies betraf vor allem die Untersuchung transparenter Objekte von BASF für Bilddatentests. Dies führte dazu, dass die eigentlich in unserem Arbeitsplan vorgesehenen Tätigkeiten demgegenüber in den Hintergrund traten.

Dass Spheron von Ximea, einem potenziellen Konkurrenten von Basler, aus der Insolvenz heraus übernommen wurde, führte im weiteren Verlauf zu Klärungsbedarf bezüglich der Zusammenarbeit und dem Teilen von Marktinformationen. Da Anfang 2024 auch der Projektkoordinator Gestalt Robotics insolvent wurde und über längere Zeit für

uns nicht erreichbar war, beschlossen wir, das Projekt intern zu pausieren, bis geklärt werden konnte, ob und wie es weitergeführt werden würde.

Auch in unserem Hause gab es wirtschaftliche Herausforderungen und die Basler AG hat 2023 aufgrund der globalen Wirtschaftsschwäche Maßnahmen zur Kostenoptimierung ergriffen. Dies führte zu einer Reduzierung des Personalbestands und zur Einführung von Kurzarbeit, was die verfügbare Arbeitszeit und die Möglichkeit zur Mehrarbeit einschränkte.

Die wirtschaftlichen und personellen Einschränkungen führten dazu, dass Basler das Projekt zum 31. März 2024 beenden musste. Die geplanten Aktivitäten zur Marktevaluierung konnten nicht angegangen werden, und es wurde entschieden, die Mitarbeit an dem Forschungsvorhaben zu beenden.

3 Wesentliche Ergebnisse (sowie ggf. Zusammenarbeit mit anderen Forschungseinrichtungen)

- Bildsensorik: Standard VIS-Sensoren wurden erfolgreich eingesetzt, um unter Zuhilfenahme geeigneter Beleuchtung transparente Objekte in 2D-Bildern darzustellen, was die geplante Nutzung teurerer und komplexerer Sensorik infrage stellte.
- Segmentierung: Die Segmentierung der transparenten Objekte in den 2D-Bildern aus VIS-Sensorik wurde mit klassischen Bildverarbeitungsmethoden erfolgreich durchgeführt. Zudem wurde mit dem KI-Modell „Segment Anything“ der Firma Meta nachgewiesen, dass AI-gestützte Methoden hier ebenfalls prinzipiell funktionieren.
- Posenbestimmung: Die Posenbestimmung der transparenten Objekte in den 2D-Bildern wurde erfolgreich durchgeführt, was eine Voraussetzung für die präzise Greifpunktbestimmung für den Roboterarm ist.
- Machbarkeit: Die Machbarkeitsstudien bestätigten, dass die Aufgabenstellung mit einem einzelnen VIS-Sensor unter bestimmten Bedingungen erfolgreich gelöst werden kann.
- Erkennung: Die Erkennung der transparenten Objekte funktionierte unter Entwicklungslaborbedingungen robust, insbesondere durch die Anpassung der Bildaufnahmeparameter wie Beleuchtung und Hintergrund.
- Beleuchtung: Die Beleuchtungstechniken wurden optimiert, um die Kanten der transparenten Objekte im Bild hervorzuheben und eine bessere Erkennbarkeit zu gewährleisten.
- Sensorfusion: Die Notwendigkeit für eine Sensorfusion wurde reduziert, da bereits mit einem einzelnen VIS-Sensor zufriedenstellende Ergebnisse erzielt werden konnten.
- Testumgebung: Die Tests und Experimente wurden unter Entwicklungslaborbedingungen durchgeführt, was half, die Ergebnisse unter kontrollierten und wiederholbaren Bedingungen zu validieren.

- - -

Sachbericht zum Verwendungsnachweis

Teil II: Ausführlicher Abschlussbericht

Verbundprojekt

Multimodale Erfassung des Normalenvektorfelds Transparenter Objekte für Roboter
MENTOR

Teilprojekt

Multimodale Bilderkennung von transparenten und semitransparenten Objekten für den industriellen Einsatz

Förderkennzeichen

13N16522

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Laufzeit

01.12.2022 - 31.03.2024

Zuwendungsempfänger

Basler AG
An der Strusbek 60-62
22926 Ahrensburg



Projektleiter

Herr Dr. Ralf Lay (bis 11/2023)
Herr Dr. Jörg Kunze (ab 12/2023)

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des [Bundesministeriums für Bildung und Forschung](#) unter dem Förderkennzeichen 13N16522 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Inhaltsverzeichnis

	<u>Seite</u>
1 Durchgeführte Arbeiten im Vergleich zur ursprünglichen Planung	1
2 Erzielte Ergebnisse	2
3 Wichtigste Positionen des zahlenmäßigen Nachweises	12
4 Notwendigkeit/Angemessenheit der geleisteten Arbeiten	13
5 Voraussichtlicher Nutzen / Verwertbarkeit der Ergebnisse	13
6 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen	13
7 Erfolgte oder geplante Veröffentlichungen des Ergebnisses	13

1 Durchgeführte Arbeiten im Vergleich zur ursprünglichen Planung

Ziel des Verbunds war die Entwicklung eines Systems, mit dem die zuverlässige (Erfolgsquote >98%) Bestimmung der Pose (= Position und Orientierung) eines Gegenstandes aus Glas („Glaskörper“) möglich ist, so dass eine Leitebene damit den Greifpunkt eines Roboters am Glaskörper ermitteln und der Roboter den Glaskörper greifen kann.

Eigentliches Ziel in unserem Teilprojekt war die Realisierung variabel konfigurierbarer Sensorik und flexibler Algorithmik zur multimodalen, bildbasierten Erkennung transparenter Objekte mit flexibel konfigurierbaren Sensor- und Systemsetups und die Entwicklung flexibel konfigurierbarer Software zur schnellen Erfassung und Vorverarbeitung, insbesondere zur Reduktion der großen Bilddatenmengen für die Übertragung an die Leitebene. Dazu sollten in einer Definitionsphase Kenngrößen und Ziele für die Systemleistung festgelegt und die Anforderungen an die Laborautomation sowie weitere Anwendungen analysiert werden. Zeitgleich waren theoretische Vorarbeiten zur Schnittstellenentwicklung geplant, um ein reibungsloses Zusammenspiel der Systemkomponenten sicherzustellen. Anschließend sollte der technische Sensorkopf ausgelegt werden. Dabei waren auch Auswahl und Integration der verschiedenen Sensoren geplant, wobei thermische, mechanische und elektrische Parameter berücksichtigt werden sollten. Die Analyse und Auswahl der Imaging-Sensoren sowie die Vorbereitung der Algorithmen-Entwicklung bilden dabei einen zentralen Schwerpunkt. Anschließend sollten Algorithmen zur Erkennung transparenter Objekte durch Kombination klassischer und KI-basierter Verfahren zur Verarbeitung multimodaler Signale realisiert werden. Ziel der Arbeiten war ein Konzept zur Erzeugung hybrider Lösungen und die Realisierung eines Software-Minimaldemonstrator auf Basis der Basler-Software Pylon. Damit sollte die Erprobung der theoretischen Ergebnisse in realen Anwendungen durchgeführt werden.

Nach der frühzeitigen Insolvenz des Projektpartners Spheron im ersten Halbjahr 2023 übernahm Basler in Absprache mit den Projektpartnern zunächst einen Teil von dessen Arbeitspaketen, die von zentraler Bedeutung für das Gesamtvorhaben waren. Vor allem betraf dies die Bildaufnahme von transparenten Objekten von BASF und optischen Linsen von Dioptics und deren Auswertung.

- Dadurch konnten wir die bei uns geplanten Arbeitspakete 1 bis 3 (Spezifikationen für unsere Anwendungsbereiche, Schnittstellen unserer Hardware zum System, Testverfahren) nur teilweise bearbeiten.
- Die Arbeiten des AP 4 (Konzept für Sensorkopf) konnten nicht durchgeführt werden, da hierzu die Zusammenarbeit mit Spheron (Entwicklung des Sensorkopfs) nötig gewesen wäre.
- Die Arbeiten der APs 5 und 6 wurden wie geplant durchgeführt, aber basierend auf den Daten und anhand von Glasgeräten der BASF und Linsen von Dioptric. Damit lag der Schwerpunkt außerhalb unseres eigentlichen Projektfokus. Dies hat aber dazu beigetragen, dass das Projekt nach dem (vorübergehendem) Ausscheiden von Spheron überhaupt weitergehen konnte.

Der Umfang der Arbeiten bei Basler war niedriger als geplant und die inhaltliche Ausrichtung geändert. Die bereits genannten und die nachstehenden Rahmenbedingungen führten auch dazu, dass die geplanten Aktivitäten zur Marktevaluierung (Transfermärkte und Transferanwendungen) nicht angegangen werden konnten.

Im Jahr 2023 hat die Basler AG aufgrund der globalen Wirtschaftsschwäche umfangreiche Maßnahmen zur Kostenoptimierung ergriffen. Dies führte zu einer Reduzierung des Personalbestands und zur Einführung von Kurzarbeit. Der initial geplante Einsatz zusätzlicher Mitarbeiter für das vorliegende Projekt, der zur Durchführung unserer eigenen Arbeiten über die übernommenen Aufgaben von Spheron hinaus notwendig gewesen wäre, war dadurch nicht mehr möglich.

Die wirtschaftlichen und personellen Einschränkungen und die fehlende Aussicht, die geplanten Arbeiten erfolgreich durchführen zu können, führten dazu, dass Basler das Projekt zum 31. März

2024 vorzeitig beenden musste. Die Arbeiten des Arbeitspakets 7 sowie der anschließenden APs wurden nicht mehr begonnen. Die Ergebnisse der durchgeführten Arbeiten sind im folgenden Abschnitt beschrieben.

2 Erzielte Ergebnisse

Basler hat Bilddaten von transparenten Objekten von BASF und optischen Linsen von Dioptics mit eigenen technischen Mitteln (Kameras, Beleuchtung, ...) erzeugt und ausgewertet. Dabei wurde Wert auf eine sinnvolle Varianz an Bildsensoren gelegt. Ziel war die Klärung der Machbarkeit der Greifpunkterkennung für die transparenten Objekte und die Definition der bildgebenden multimodalen Sensorik.

Die dabei erzielten Ergebnisse wurden allen anderen Teilnehmern zur weiteren Verwendung mitgeteilt. Für Details dazu siehe unter '5 Voraussichtlicher Nutzen / Verwertbarkeit der Ergebnisse'.

Thematisch haben wir die von Spheron übernommenen Arbeiten nachstehend unseren APs 5 und 6 zugeordnet und die vorab durchgeführten Arbeiten den geplanten APs 1 bis 3. In den folgenden Abschnitten sind die Ergebnisse entsprechen dieser Zuordnung dargestellt.

AP1 (Definition der Specs für den Anwendungsbereich bei Basler):

Die Evaluierung der im Verbund geplanten Anwendungen und Anforderungen ist bei Basler aufgrund der oben genannten Rahmenbedingungen und abweichend zur initialen Aufgabenstellung für die transparenten Objekte von BASF erfolgt. Diese Anwendung war für das Verbundvorhaben von grundlegender Bedeutung. Weitere Anwendungsbereiche und damit wesentliche Themen aus dem AP1 für Basler wurden dementsprechend zurückgestellt.

Mit BASF wurden Gestaltungsoptionen für die Versuche bei Basler abgeglichen, wie sie unter den folgenden APs beschrieben sind (Hintergrund, Beleuchtung, ...).

AP2 (Theoretische Vorarbeiten zu den Schnittstellen im System)

In diesem AP stand die spätere Implementierung auf Embedded-Hardware im Fokus. Für die Anwendung im Bereich Embedded sind Kameramodule mit CIS-Interface die erste Wahl und nicht komplette Kameras mit z.B. USB- oder GigE-Interface.

Moderne SoC's für Embedded Processing haben bis zu 12 Lanes MIPI CSI-2 mit denen z.B. 6 Kameramodule an jeweils 2 Lanes betrieben werden können (Nvidia Jetson TX1, TX2, ...). Damit ist eher das Processing der Bilddaten ein Engpass bezüglich der Datenrate und weniger das Kamerainterface. Um die Situation beurteilen zu können, ist es wichtig zu wissen, wie viele Bildsensoren welcher Art (Auflösung, Pixeldatenformat) auszuwerten sind. Zudem ist die erforderliche Bildrate relevant. Nicht zuletzt spielt auch die Antwort auf die Frage, wie die Auswertung erfolgt, eine entscheidende Rolle. Das Ergebnis einer ersten Abschätzung für das Zusammenspiel der Sensorik, Algorithmik und Hardware und der erreichbaren Datenraten konnte hier folglich nur lauten: Die Chance, dass es funktioniert, ist gegeben, es hängt aber von den Details ab, die an dieser Stelle noch nicht bekannt waren.

Die späteren Ergebnisse von Basler, dass man unter bestimmten Bedingungen mit einem einzelnen VIS-Bildsensor (,VIS' steht für ,visible' light) schon weit kommen kann, führen zu einer positiven Abschätzung der grundsätzlichen Machbarkeitsfrage für die bildbasierte Erkennung der transparenten Objekte. Diese Aussage ersetzt nicht die Notwendigkeit für die Durchführung von weiteren Tests bzw. die Erstellung von Demonstratoren mit denen das Konzept auch unter realen Bedingungen geprüft und ggf. optimiert werden kann.

AP3 (Festlegung notwendiger Leistungskriterien und Testverfahren aus Sicht von Basler)

Eine Analyse der benötigten Testverfahren und Prüfaufbauten der Objekte von BASF führte zu den nachstehend beschriebenen Experimenten und Ergebnissen unter Entwicklungslaborbedingungen. Aus anderen Projekten bei Basler verfügbare Tools, Aufbauten und Testmittel konnten mit Anpassungen dafür sinnvoll eingesetzt werden.

AP4 (Überführung der Spezifikationen in ein konkretes technisches Konzept für den Sensorkopf):

Da Spheron insolvent gegangen war, gab es von dort keine Vorergebnisse, auf denen wir für ein konkretes technisches Konzept des Sensorkopfes hätten aufbauen können. Basler hat sich hier selbst um die Aufgaben von Spheron gekümmert, die Objekte von BASF untersucht und insbesondere hinsichtlich AP5 (Sensorik) und AP6 (Algorithmik) eine konkrete Lösungsoption herausgearbeitet.

AP5 (Analyse der Anforderungen und Auswahl der Imaging-Sensoren Vorbereitung der Entwicklungsarbeiten zur Algorithmik)

Bildaufnahme / Sensorvergleich

Im AP 5 wurden (abweichend von der eigentlichen Planung) zunächst transparente Objekte für den assoziierten Partner BASF (Abbildung 1: Transparente Objekte von BASF) sowie optische Linsen des Projektpartners Dioptics (Abbildung 2: Optische Linsen von Dioptics) analysiert.



Abbildung 1: Transparente Objekte von BASF

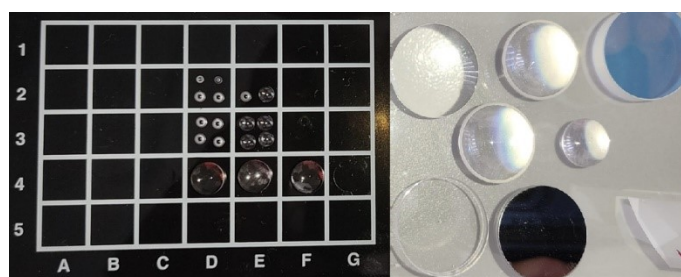


Abbildung 2: Optische Linsen von Dioptics

Bei Hellma Analytics waren Details zu spektralen Transmissionskurven transparenter Materialien zu finden (Abbildung 3). Diese ermöglichten es, die Chancen einer Bildsensorik für die gegebene Vielfalt der zu testenden Objekte einzuordnen. Quartz-Glas ist z.B. in einem weiten

Spektralbereich von 175nm bis 4000nm transparent. Für die anderen Gläser gilt dies im Bereich 300nm bis 2700nm.

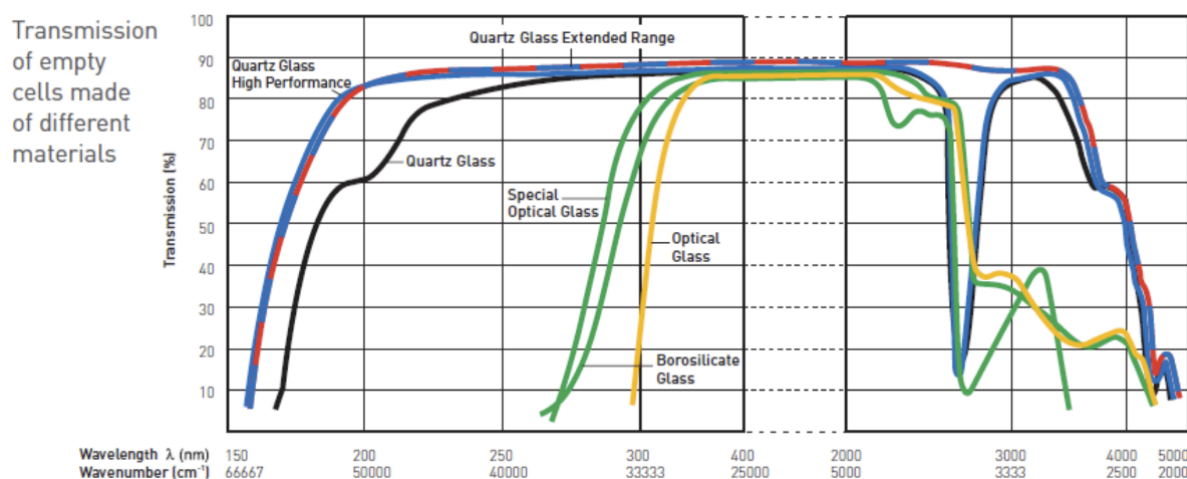


Abbildung 3: Spektrale Transmissionskurven transparenter Materialien (Quelle: Hellma Analytics)

Für den Fall einer nichttransparenten Darstellung der gläsernen Objekte galt die Machbarkeit der Greifpunktbestimmung nach Stand der Technik grundsätzlich als gegeben. Dies wurde beispielhaft mit einer Optris Kamera vom Typ PI640i nachvollzogen. Im Bereich 5500nm waren hier alle Testobjekte ‚kooperativ‘, d.h. nicht transparent (opak). Diese Kameras sind aber teuer und sie passten nicht zu dem Ziel, preiswerte Systeme zu bekommen.

Angeichts der Aufgabe des Projektes, bei moderaten Kosten der Sensorik transparente Objekte im Bild erkennen und deren Greifpunkte bestimmen zu können, wurde das Konzept „multimodale Bilderfassung“ untersucht, um damit zu hinreichend guter Funktionalität zu kommen.

Es wurden Methoden auf Basis von 2D-Bildsensoren untersucht. Die Variation der Modalität bestand u.a. im spektralen Erfassungsbereich der Bildsensorik. Dabei wurde vorzugsweise der Bereich getestet, in dem alle Gläser transparent sind. Zudem wurden ein LWIR-Sensor und ein Polarisationsensor herangezogen.

In den nachstehenden Bildern war hinter dem transparenten Objekt eine kleine Glühlampe angebracht. Sie hat als Thermostrahler einen sehr breitem Wellenlängenbereich. Damit ist die tatsächliche Transparenz bzw. die „Durchsichtigkeit“ des Objektes belegt.

Die Versuche erfolgten zunächst unter Entwicklungslaborbedingungen und noch nicht mit realen Umgebungsbedingungen, die für die Roboter im Bereich Lab-Automation gelten. Die Szene wurden für die Versuche mit einem definierten Hintergrund versehen (oft diffus schwarz, was auch einem zulässigen Hintergrund im späteren realen Umfeld entspricht). Ziel dabei war es, den Hintergrund im Bild homogen, reflexionsarm und im validen Grauwertbereich zu halten (d.h. z.B. nicht überstrahlt).

Die Bilder wurden in einer professionellen Lichtbox aus unserem hausinternen Fotostudio aufgenommen. Als Beleuchtung diente eine ringförmige LED-Lichtquelle.

Problematisch waren Spiegellungen auf dem Glas. Man konnte z.B. teilweise die Kamera sehen. Dies kann prinzipiell durch Abdunklung der Bildaufnahmeeinheit unterbunden werden, was auch für den späteren Roboterarm prinzipiell möglich ist. Es wäre aber nicht hinreichend: Auf den gläsernen Objekten wird die gesamte rückseitige Szene gespiegelt. Durch die runde Form der Objekte wirkt ein weiter ‚Sichtwinkel‘. Das alles abzdunkeln ist quasi nicht möglich. Bei transparenten Objekten kommt erschwerend hinzu, dass mit ihrer Durchsichtigkeit auch deren Hintergrund in den Bildern mit dargestellt wird. Beides führte zu potentiell undefinierten Bereichen auf den Objekten in den Bildern, die die Objekterkennung grundsätzlich erschweren. Gelingt es aber, die Außenkanten der transparenten Objekte im Bild gut sichtbar zu machen und die algorithmische Auswertung nur dort anzusetzen, wäre es möglich, die kritischen, potentiell

undefinierten Bereiche ‚auszublenden‘. Auch für die Greifpunkterkennung wäre es vorteilhaft, die Außenkanten des jeweiligen Objektes im Bild definiert sichtbar zu haben.

Zur Sichtbarmachung der Außenkanten wurde die Belichtung variiert und Bildmaterial in Belichtungsreihen generiert und ausgewertet. Ziel war es, herauszufinden, ob die transparenten Objekte in Überbelichtung an den Außenkanten im Kontrast zum dunklen Hintergrund darstellbar sind. Dies konnte bestätigt werden, wie es beispielhaft in Abbildung 4 und Abbildung 15 zu sehen ist.



Abbildung 4: Beispiel für die Überbelichtung eines Objektes vor dunklem Hintergrund

Es folgen Beispielaufnahmen mit verschiedenen Bildsensorvarianten (Abbildung 5 bis Abbildung 9). Dabei ist zu beachten, dass noch keine Überbelichtung zur Hervorhebung der Objektkanten angewendet wurde. Die Bilder dienen zunächst dem qualitativen Vergleich der Sensorvarianten.



**Abbildung 5: Standard-VIS-Sensor monochromatisch z.B. in Kamera acA1920-48gm.
Glas (links) und Quarzglas (rechts)**



Abbildung 6: VIS-SWIR-Sensor in Kamera a2A1280-80gmSWIR. Glas



Abbildung 7: SWIR-Sensor in a2A1280-80gmSWIR mit Filter Thorlabs FEL1000 (VIS-SWIR-Kamera mit VIS-Filter bei Cut-On Wavelength 1000nm). Glas

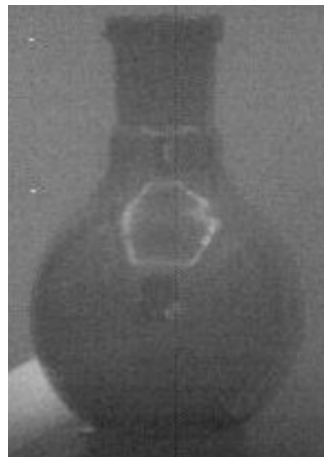


Abbildung 8: LWIR in Kamera A320-50g-LW-19mm (Wellenlängen größer 4500nm). Glas



Abbildung 9: Polarisationsensor in Kamera: acA2440-75ucPOL. Glas, Quarzglas und optische Linse

Ein UV-Bildsensor wurde nicht getestet. Da sich in realen Laborumgebungen auch Menschen bewegen, könnte eine UV-Beleuchtung kritisch sein.

Es folgen die Ergebnisse aus dem Vergleich der Bildsensorvarianten.

Die Bildaufnahme mit Sensoren von VIS bis SWIR zeigten keine Vorteile hin zu größeren Wellenlängen. D.h., ein kostengünstiger monochromatischer VIS-Sensor kann ohne Nachteile eingesetzt werden. Bedingung ist ggf. eine Überbelichtung, um die Kanten der gläsernen Objekte gegenüber dem Hintergrund besser hervorzuheben.

Beim LWIR-Sensor mit Wellenlängen größer als 4500nm wurden die Objekte opak. Die Auflösung dieser Sensorik war allerdings deutlich eingeschränkt und die Kontraste im Bild fielen schwach aus. LWIR-Kameras sind komplexer und teurer als z.B. VIS-Kameras, die sich hier somit als die bessere Wahl darstellten.

Ein Polarisationsensor zeigte gegenüber einem VIS-Sensor keine wesentlichen Vorteile in der ‚Erkennbarkeit‘ der transparenten Objekte aus Glas und Quarzglas. Es bestand die Erwartung, mit einem Sensor, bei dem jeder Pixel einer 4er Gruppe einen eigenen, anders gedrehten Polfilter besitzt, weniger Reflexionen zu bekommen. Das hat sich aber nicht bestätigt.

Der Vergleich zeigte, dass ein Standard-VIS-Sensor insgesamt die beste Wahl darstellt. Eine latente Überbelichtung kann dazu dienen, die Objektkanten im Bild deutlicher hervorzuheben, was in Abbildung 5 noch nicht angewendet war.

AP6 Grundlagen für Algorithmen zur Erkennung transparenter Objekte

Algorithmik

Die Bildverarbeitung für die transparenten Objekte in den aufgenommenen 2D Bildern unterteilt sich in die Arbeitsschritte:

- 1) Segmentierung
- 2) Posenbestimmung
- 3) Greifpunktbestimmung

Es bedeutet quasi eine 3D-Rekonstruktion der transparenten Objekte in typischen Laborbedingungen aus 2D-Bilddaten. Die Option des Sensor-Fusion bringt die Anforderung für Image-Fusion bzw. Data Fusion in die Bildauswertung ein.

Wir haben dazu auf konzeptioneller Ebene die Optionen eines AI-gestützten **End2End-Trainings** bewertet. Hiermit wären potentiell alle 3 Teilschritte der Bildauswertung auch unter Berücksichtigung von Sensor-Fusion/Data-Fusion in einem AI-Modell lösbar.

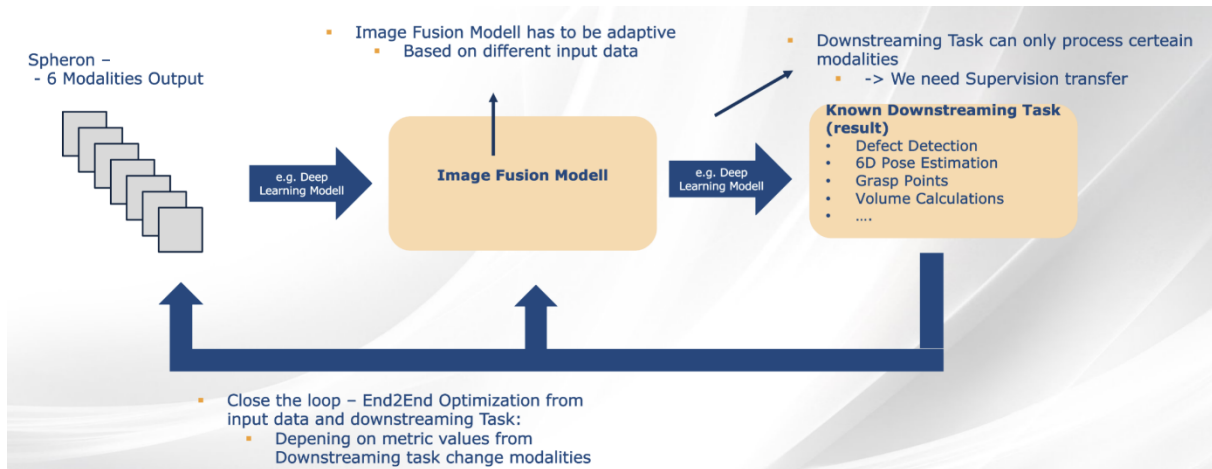


Abbildung 10: Bewertung eines End2End-Trainings als Option bei Image Fusion

Die Vorteile eines End2End-Trainings bzw. Modells sind:

- Verwendung von Standard-Downstream Tasks und bekannter Netze möglich
- Das Deep-Learning Modell organisiert sich beim Training selbst
- Explizite bzw. manuell gestaltete Netzkonzepte können vermieden und Entwicklungszeit gespart werden

Dem stehen folgende Herausforderungen eines End2End-Trainings gegenüber:

- Transfer Learning für die Downstreaming Tasks
- Adaptives Image Fusion Modell (abhängig von den Input-Daten bzw. der Sensor-Konfiguration zur Bildaufnahme)

Für **Data-Fusion** bei multimodalem Sensoreinsatz fanden sich (ebenfalls in früher konzeptioneller Projektphase) in externen Quellen mehrere mögliche Ansätze:

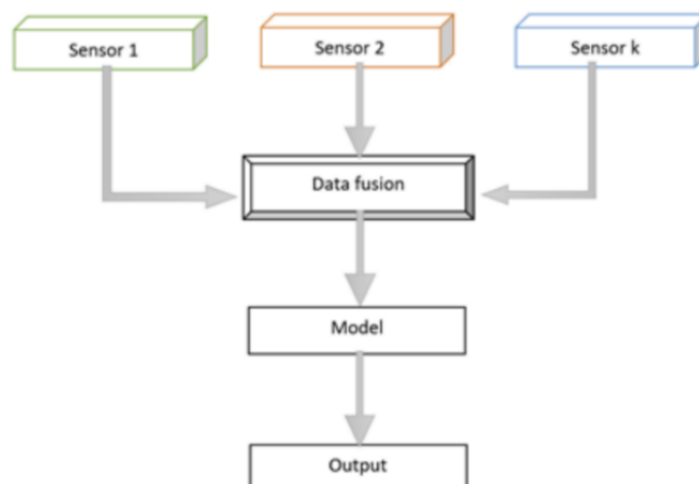


Abbildung 11: Early fusion or data-level fusion

(<https://medium.com/haileleol-tibebu/data-fusion-78e68e65b2d1>)

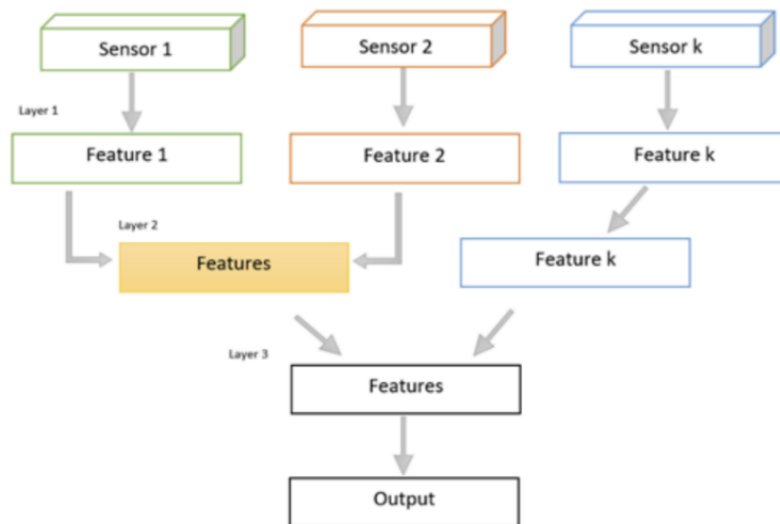


Abbildung 12: Intermediate fusion

<https://medium.com/haileleol-tibebu/data-fusion-78e68e65b2d1>

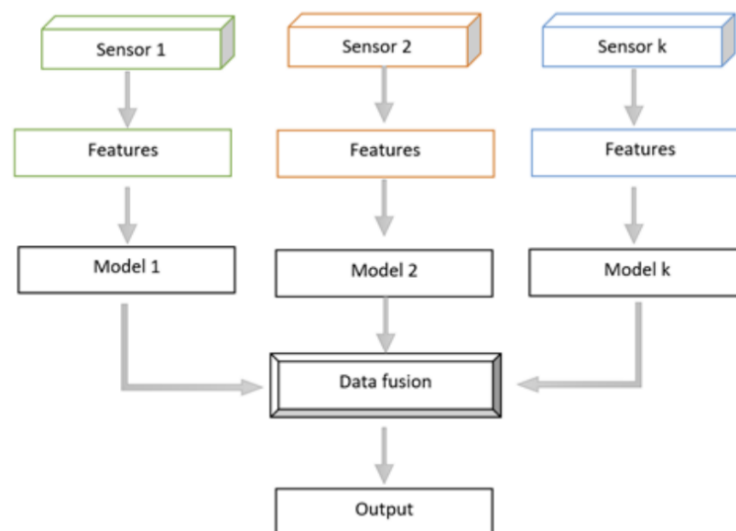


Abbildung 13: Late fusion or decision fusion

<https://medium.com/haileleol-tibebu/data-fusion-78e68e65b2d1>

Die Ergebnisse der Machbarkeitsuntersuchungen für die Bildaufnahme zu AP5 von Basler (s.o.) zeigten, dass man unter bestimmten Bedingungen mit einem einzelnen VIS-Bildsensor schon weit kommen kann. Die Notwendigkeit für multimodale Bildgebung würde damit vermieden werden und eine weitere Betrachtung von multimodalem Image- bzw. Data Fusion wäre nicht mehr erforderlich.

Praktische Machbarkeitsuntersuchungen

Die bisherigen Ergebnisse führten zu dem Ansatz, die Aufgabenstellung zunächst mit einem einzelnen Standard-2D-VIS-Sensor anzugehen und zu erproben, wie weit sich dies auch hinsichtlich der Algorithmik als machbar darstellt.

Der Schwerpunkt lag zunächst auf der Segmentierung und der Posenbestimmung der transparenten Objekte. Die Idee war, einen Roboterarm für die automatisierte Bildaufnahme zu nutzen. Dafür wurde ein bei Basler bereits vorhandener Roboterarm mit einem Halter für eine Kamera ausgerüstet. Mit angepassten Python-Skripten wurde der Roboterarm automatisch in definierte Positionen gefahren, um Bildaufnahmen zu tätigen. Bei Kenntnis der genauen Position

der Kamera zum Objekt erhält man ‚Ground Truth‘ - Bilder. Diese können z.B. eingesetzt werden, um über Augmentation vollständige Datensätze für ein Training von AI-Netzen zur Bildauswertung zu erzeugen.

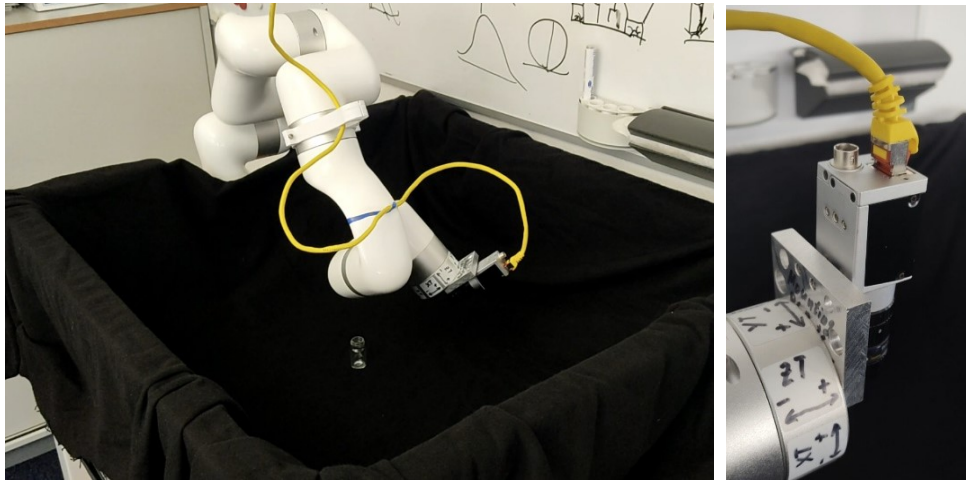


Abbildung 14: Roboterarm zur Bildaufnahme (z.B. im Halbkugelkreis um das Objekt)



Abbildung 15: VIS-Sensor: Ground-Truth-Bilder mittels Roboterarm und einer VIS-Kamera aufgenommen. Im Bild rechts mit Überbelichtung, was der Erkennung der Kontur der transparenten Objekte dient

Der Plan war, den Roboteraufbau neben der Bildaufnahme für das Training von AI-Modellen auch für einen automatisierten, umfänglichen Funktionsnachweis der Bildaufnahme plus der Algorithmik zur Bildauswertung zu nutzen. Vorbereitend wurde dazu im Rahmen des Förderprojektes bei Basler ein Greifer für den gegebenen Roboterarm beschafft und ein erster Test für das Greifen eines Gegenstandes durchgeführt (Abbildung 16).

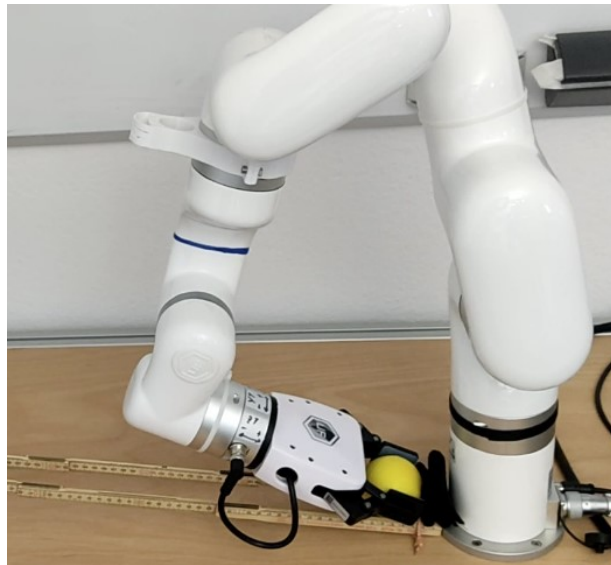


**Abbildung 16: Greifen von Gegenständen mittels Roboterarm
(hier: Ball zyklisch greifen und auf eine sehr rudimentär
simulierte Rampe legen)**

Die Objektsegmentierung wurde in einem ersten Schritt mit dem im Internet verfügbaren Tool ‚Segment Anything‘ der Firma Meta (<https://segment-anything.com/>) getestet.

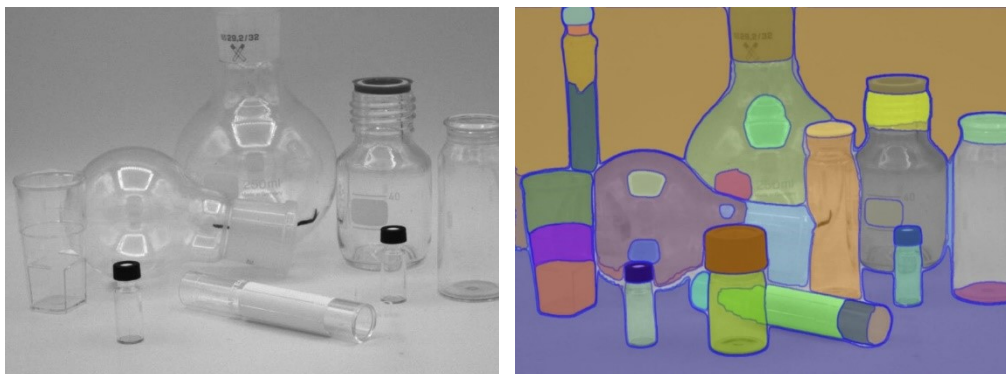


Abbildung 17: Segmentierung mittels Segment Anything – Originalbild und Ergebnisbild

Damit war die Segmentierung von sich teilweise überlappenden, durchsichtigen, reflektierenden Gläsern auch vor einem hellen Hintergrund und in nur einer Aufnahme mit Standard-Bildsensorik (2D, VIS, monochrom) gut möglich.

In einem weiteren Schritt wurden das Bildaufnahme-Setup und die Algorithmik zur Objekterkennung aus einem separaten Entwicklungsprojekt von Basler herangezogen. Es war mit einer Standard-2D-VIS-Kamera ausgestattet. Die Algorithmik basierte bei Nutzung der kommerziell verfügbaren Bildverarbeitungsbibliothek HALCON von der MVTec Software GmbH auf rein klassischer Bildverarbeitung. Die Algorithmik war in diesem System primär für nichttransparente Objekte ausgelegt. Sie funktionierte nach sinnvoller Parametrierung der Bildaufnahme inkl. Beleuchtung dennoch auf Anhieb gut für die Erkennung der hier gegebenen transparenten Objekte, d.h., die Segmentierung und die Posenbestimmung war möglich. Dabei galten die folgenden Voraussetzungen:

- Relativ homogener Bildhintergrund
- Außenkanten der Objekte im Bild erkennbar (ggf. Überbelichtung nutzen)
- Speziell für die Posenerkennung
 - Die 3D-CAD-Daten der Objekte liegen vor
 - Das Bildaufnahmesystem ist intrinsisch kalibriert (geometrisch)

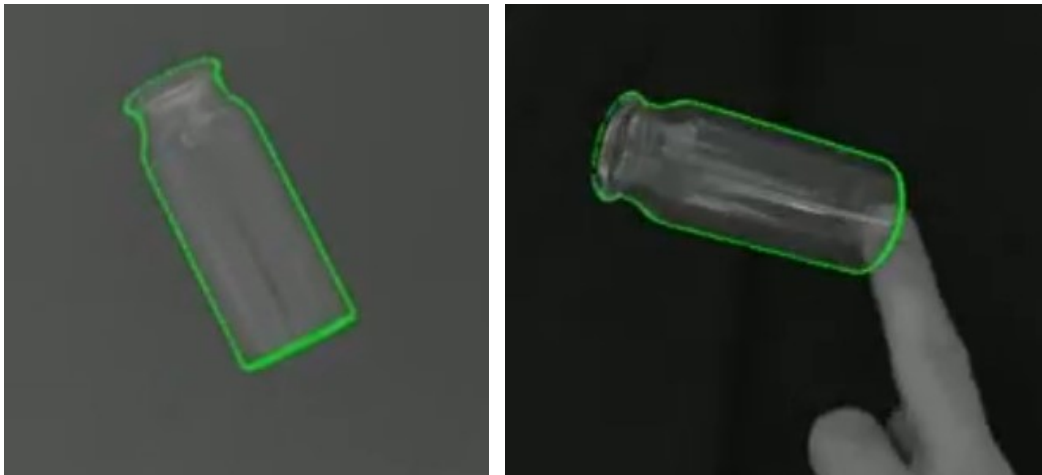


Abbildung 18: Objekterkennung und Posenbestimmung aus einem Parallelprojekt auf Basis von HALCON und 3D-CAD-Daten der Objekte

Weitergehende Tests für das Ziel der Evaluierung der Greifpunkterkennung aus den Daten der Objektsegmentierung und Posenbestimmung transparenter Objekte in 2D-Bildern sind seitens Basler nicht erfolgt.

Zusammenfassung der erzielten Ergebnisse / Fazit:

Die Segmentierung der transparenten Objekte (von BASF) funktionierte unter definierten Entwicklungslaborbedingungen gut. Wir konnten mit einer Standard-VIS-Kamera die Objekte und deren Kanten im 2D-Bild finden. Dies funktionierte auch mit klassischer Bildverarbeitung und gängigen Bildverarbeitungs-Bibliotheken wie HALCON. Es konnte gezeigt werden, dass AI-Lösungen nutzbar sind. Neben dem schnellen Test mit Segment Anything wurde ein erster Aufbau zur Gewinnung von Ground Truth-Bildern erzeugt.

Das genannte System aus einem Parallelprojekt auf Basis von HALCON war bei Kenntnis der 3D-CAD-Daten der Objekte in der Lage, auch die Posenbestimmung zu leisten. Damit konnte die prinzipielle Machbarkeit der Posenbestimmung für transparente Objekte unter den genannten Testbedingungen für 2D-Bilddaten aus (nur) einem Standard-VIS-Bildsensor gezeigt werden.

Die Ergebnisse zeigen, dass die Erkennung der transparenten Objekte in 2D Bildern einfacher zu lösen ist als initial angenommen. Unter Entwicklungslaborbedingungen kommt man mit einer Standard-Bildsensorik aus nur einem VIS-Sensor + Beleuchtung und klassischer Bildverarbeitung bei Kenntnis der CAD-Daten der zu greifenden Objekte schon recht weit. Eine AI-basierte Bildauswertung bei einem (einzelnen) VIS-Bildsensor ist eine weitere, vielversprechende Option. Das Ergebnis kann vermutlich auch auf reale Einsatzbedingungen übertragen werden, wenn dabei gewisse Gestaltungsoptionen bestehen. Mit BASF wurde explizit abgeglichen, derartige Optionen (Hintergrund, Beleuchtung, ...) in der Versuchsphase verwenden zu können.

Die konzeptseitig angedachten Optionen für Image- bzw. Data-Fusion und ein End2End-Training und wurden von Basler aufgrund der bereits erreichten Ergebnisse nicht weiter untersucht.

Nach der Segmentierung im 2D-Bild und der Posenbestimmung der Objekte stellt die Bestimmung der Greifpunkte für den Roboterarm den finalen Schritt der Bildauswertung dar. Dieser wurde seitens Basler aufgrund der eingangs genannten internen Rahmenbedingungen nicht weiter evaluiert.

3 Wichtigste Positionen des zahlenmäßigen Nachweises

Im Wesentlichen sind während der Projektlaufzeit Personalkosten entstanden. Der verkürzten Laufzeit entsprechend fielen ca. 33% der geplanten Aufwendungen an. Auch Reisekosten und Abschreibungen entsprechen der verkürzten Laufzeit.

4 Notwendigkeit/Angemessenheit der geleisteten Arbeiten

Die geleisteten Arbeiten entsprachen zwar nur zum Teil der Planung, alle Abweichungen wurden aber mit den Projektpartnern und dem Projektträger abgestimmt. Wie dem Ergebnisteil des Berichts entnommen werden kann, wurden dabei wesentliche Ergebnisse erzielt, die von den Partnern benötigt wurden. Dafür waren zwar teils umfangreiche Versuchsreihen notwendig, die aber in jedem Fall angesichts des Ergebnisses aus unserer Sicht angemessen erscheinen und für die Zielerreichung notwendig waren.

Die zwischen den Partnern notwendigen Abstimmungen haben zwar für sich gesehen ebenfalls Kapazitäten gebunden, durch die entstandenen Synergieeffekte aber signifikant zu einer effizienten Projektabwicklung beigetragen.

Die im Projekt eingesetzten Mittel für Verbrauchsmaterialien waren unabdingbar für die Durchführung der Versuche und im angefallenen Umfang unverzichtbar für die Arbeiten.

5 Voraussichtlicher Nutzen / Verwertbarkeit der Ergebnisse

Zu Beginn des Projekts bestand die Perspektive, in neue Märkte vorzudringen, wirtschaftlichen Nutzen zu erzielen und neue Arbeitsplätze zu schaffen. Im Laufe des Projekts haben sich die Randbedingungen seitens Basler aber intern aus wirtschaftlichen Gründen und extern durch Insolvenzen bei Verbundpartnern und anschließenden Übernahmen verändert, was dazu geführt hat, dass Basler aus dem Projekt ausgestiegen ist, um für alle Seiten keine weiteren Kosten zu erzeugen.

Eine Verwendung der Projektergebnisse kann seitens Basler mit Bezug auf die Rahmenbedingungen nicht erfolgen. Nutzungen der Ergebnisse durch die Projektpartner werden von diesen im weiteren Verlauf der (fortgeführten) Teilprojekte untersucht.

Die von Basler erzielten Projektergebnisse wurden allen anderen Teilnehmern zur weiteren Verwendung mitgeteilt. Dies im Zeitraum März 2023 bis August 2023 im Zuge mehrerer gemeinsamer Online-Meetings und im August 2023 durch Hochladen der gesammelten Daten und Bilder in den gemeinsamen Teams-Ordner. Sollte es darüber hinaus Anfragen von den Partnern geben, erhalten diese weitere Auskunft.

6 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit sind die nachstehenden Success Stories zu der auch von Basler genutzten Bildverarbeitungsbibliothek HALCON von der MVTec Software GmbH bekannt geworden. Wie beschrieben, konnte Basler mit HALCON die Machbarkeit der Segmentierung und Posenbestimmung der gläsernen Objekte unter Nutzung eines einzelnen 2D-VIS-Bildsensors und klassischer Bildverarbeitung demonstrieren. Die folgenden Quellen konnten daraufhin aus aktiver Internetrecherche gefunden werden. Sie bestätigen die Ergebnisse von Basler. Relevante Stichworte sind hier z.B.: „Shape-Based 3D Matching“, „Parts made of glass“.

<https://www.mvtec.com/application-areas/success-stories/pick-and-place-solution-in-the-medical-field>

<https://www.mvtec.com/application-areas/success-stories/a-fully-automated-production-system-for-higher-throughput-and-improved-quality>

7 Erfolgte oder geplante Veröffentlichungen des Ergebnisses

Eine Veröffentlichung der beschriebenen Ergebnisse erfolgte nicht und ist auch wegen des vorzeitigen Projektendes nicht mehr geplant.