Schlussbericht zum Verbundprojekt HiAvA

Zuwendungsempfänger: The Captury GmbH
Förderkennzeichen: 16SV8782
Berichtszeitraum: 01.09.2021 – 31.08.2024

Dieser Schlussbericht fasst die Durchführung und die erzielten Ergebnisse des Verbundprojekts HiAvA zusammen, mit besonderem Fokus auf die Beiträge der The Captury GmbH (CAP). Das Projekt HiAvA erforschte und entwickelte skalierbare hybride Avatar-Agenten-Technologien für alltagstaugliche soziale Interaktion in XR.

Zusammenfassung

Das Projekt HiAvA hatte zum Ziel, innovative Technologien für verkörperte Multi-User-Anwendungen in sozialer VR zu erforschen und zu entwickeln. Anknüpfend an den Stand der Technik, der in Bezug auf Skalierbarkeit, Reproduktionsgüte und Komfort von Social-VR-Lösungen noch erhebliche Mängel aufwies, konzentrierte sich das Projekt auf die Entwicklung berührungsloser Trackinglösungen für Körper und Gesichter, die Verbesserung der Reproduktionsgüte durch dedizierte Rekonstruktionsverfahren und die Erhöhung der Skalierbarkeit durch die Unterstützung unterschiedlicher Endgeräte. Die The Captury GmbH war maßgeblich an der Entwicklung und Integration der Echtzeit-Tracking-Komponenten beteiligt.

Der Projektverlauf wurde in enger Zusammenarbeit der Verbundpartner The Captury GmbH (CAP), Julius-Maximilians-Universität Würzburg (JMU), Technische Universität Dortmund (TUD), Universität Bielefeld (UBI), SPREE Interactive GmbH (SI) gestaltet. Captury konzentrierte sich dabei auf das Arbeitspaket 4 "Responsives Tracking", das die Integration bestehender Tracking-Technologien sowie die Entwicklung eines neuen Mimik-Erkennungsverfahrens umfasste. Es wurden synthetische und reale Trainingsdaten für das Gesichts- und Augentracking generiert und zur Entwicklung robuster KI-Modelle genutzt.

Die wesentlichen Ergebnisse von Captury umfassen:

- **Integration von OptiTrack:** Eine nahtlose Echtzeit-Integration von OptiTrack-Kameras für Infrarot-Daten wurde entwickelt und durch technischen Support von OptiTrack unterstützt.

- **Handtracking-Erweiterung:** Das interne Skelettmodell wurde um Fingerknochen erweitert, die Trackingarchitektur für parallele KI-Modelle angepasst und das Streaming sowie automatische Retargeting für Fingerbewegungen implementiert.
- **Gesichts- und Augentracking:** Es wurden zwei separate, hochleistungsfähige Gesichtstracking-Modelle entwickelt: eines für am Kopf getragene Kameras (Head-Mounted Camera) und ein weiteres für statische Kameras (klassisches Mocap und Webcams wie in Videokonferenzen eingesetzt). Diese Modelle wurden durch umfangreiche synthetische und verfeinerte reale Trainingsdatensätze sowie spezialisierte Fixes für häufige Qualitätsprobleme (z.B. Mund- und Augenstatus) optimiert.
- **Temporale Konsistenz:** Techniken zur Verbesserung der temporalen Konsistenz des Gesichtstrackings wurden implementiert, einschließlich eines frequenzadaptiven Tiefpassfilters und eines "Wobble-Fix", der die natürliche Bewegung von Gesichtsmerkmalen besser abbildet.

Die Zusammenarbeit mit den anderen Forschungseinrichtungen war intensiv und fruchtbar, insbesondere mit der TUD bei der Entwicklung des Blendshape Retargeting und mit JMU bei der Integration in die Unity-basierte Projektplattform.

Eingehende Darstellung

1. Projektfortschritt und erzielte Ergebnisse im Vergleich zur ursprünglichen Vorhabenbeschreibung

Die The Captury GmbH war im Verbundprojekt HiAvA maßgeblich für das Arbeitspaket 4 "Responsives Tracking" verantwortlich, mit dem Ziel, berührungslose Trackinglösungen für Körper, Mimik und Augenbewegungen zu entwickeln und zu integrieren.

AP 4.1 Streaming von OptiTrack in CapturyLive (3PM):
Dieser Arbeitsschritt wurde erfolgreich abgeschlossen. Eine robuste Integration für OptiTrack-Kameras wurde entwickelt, die die Echtzeit-Verarbeitung von Infrarot-Daten in das Captury-System ermöglicht. Die Unterstützung des nativen Kalibrationsformats von OptiTrack gewährleistete eine reibungslose Integration. Dies übertraf die ursprüngliche Planung, da die Kooperation mit OptiTrack zu einem besseren Ergebnis führte als erwartet.

AP 4.2 Anbindung Handtracking (2PM):
Dieses Arbeitspaket wurde vollständig umgesetzt. Das interne Skelettmodell von Captury wurde erfolgreich um Fingerknochen erweitert. Die Trackingarchitektur wurde angepasst, um mehrere KI-Modelle (Ganzkörper- und Finger-Tracking) parallel betreiben zu können.

Streaming und automatisches Retargeting wurden ebenfalls für die Unterstützung von Fingerbewegungen optimiert.

AP 4.3 Anbindung Gesichts- & Augentracking aus iPhone, Vive Eye and Facial Tracking SDK (3PM):
In Abstimmung mit den Projektpartnern wurde entschieden, Gesichts- und Eye-Tracking-Daten direkt in die Game Engine zu streamen, um die Latenz des Gesamtsystems zu reduzieren. Die ursprünglich für die Anbindung an die Captury Software vorgesehenen Projektmonate wurden für die Entwicklung eines zweiten Gesichtstracking-Modells verwendet. Es hat sich als vorteilhaft erwiesen, separate Modelle für am Kopf getragene Kameras (HMCs) und statische Kameras zu entwickeln, was eine signifikante Verbesserung gegenüber dem ursprünglichen Plan darstellt, der eine allgemeinere Anbindung vorsah.

AP 4.4 Trainingsdaten Gesichtstracking aufnehmen (3PM):
Dieses Arbeitspaket wurde kontinuierlich über die gesamte Projektlaufzeit hinweg bearbeitet. Es wurden umfassende Validierungs- und Refinement-Datensätze sowohl für das Helmkamera- als auch für das statische Kamerasystem aufgenommen. Ein enger Loop mit AP 4.8 ermöglichte die Identifizierung und Behebung spezifischer Tracking-Probleme durch die Aufnahme gezielter Zusatzdaten (z.B. für den "Mouth Open Fix"). Die Erkenntnis, dass synthetische Daten gut auf echte Aufnahmen generalisieren, führte zu einer Verschiebung des Arbeitsaufwands zugunsten von AP 4.5.

AP 4.5 Trainingsdaten Gesichtstracking generieren (3PM):
Die Generierung synthetischer Trainingsdaten war ein Schlüsselbereich für die Verbesserung der Trackingqualität. Eine komplett überarbeitete Bild-Generierungs-Pipeline ermöglichte die Erzeugung von Millionen verschiedener Individuen in unterschiedlichen Posen, was die Varianz der Trainingsdaten drastisch verbesserte. Verbesserte Post-Processing-Verfahren erhöhten zusätzlich die Datenvarianz beim Training. Dies war ein erheblicher Fortschritt gegenüber dem ursprünglichen Plan, der sich möglicherweise stärker auf reale Daten konzentriert hätte.

AP 4.6 Blendshape Retargeting (3PM):
Dieser Arbeitsschritt wurde erfolgreich abgeschlossen. Ein Blendshape-Retargeting-Verfahren wurde entwickelt, das die Übertragung der Modellausgaben auf andere Charaktermodelle ermöglicht, auch wenn diese unterschiedliche Interpretationen der Blendshapes oder andere Blendshape-Sets verwenden. Dies ist ein wichtiger Beitrag zur Flexibilität und Anwendbarkeit der entwickelten Avatare.

AP 4.7 Reduziertes Blendshape Modell (6PM):
Initiale Experimente zeigten, dass eine zu starke Reduktion der Blendshape-Anzahl zu

einem Verlust subtiler Gesichtseffekte führte. Stattdessen wurde die Zeit in die Optimierung des Trainingsdatenerzeugungs- und Modelltrainings-Loops investiert. Dieses Arbeitspaket wurde als abgeschlossen betrachtet, da die erzielten Ergebnisse durch die Verfeinerung der anderen Arbeitspakete, insbesondere AP 4.8, die gewünschten Qualitätsziele erreichten, auch ohne ein explizit reduziertes Blendshape-Modell einzusetzen.

AP 4.8 KI Modell trainieren (5PM):
Dieser zentrale Arbeitsschritt wurde mit erheblichem Erfolg abgeschlossen. Ein enger Loop von Modelltraining und Evaluation führte zu sehr guten Ergebnissen für das Gesichts-Tracking, insbesondere bei Daten, die direkt vor dem Gesicht aufgenommen wurden (Helmkamera). Folgende Qualitätsprobleme wurden im Rahmen der Entwicklung identifiziert und erfolgreich behoben:

- **Mouth Open Fix:** Durch synthetische Constraints wurde die Tendenz des Modells, den Mund immer leicht geöffnet zu schätzen, behoben.
- **Mouth Closed Fix:** Reale Daten wurden eingesetzt, um Situationen zu beheben, in denen der Mund fälschlicherweise als geschlossen geschätzt wurde.
- **Teeth Not Closed Fix:** Synthetische Constraints korrigierten die fehlerhafte Schätzung bei gefletschten Zähnen.
- **Eyes Closed Fix:** Reale Daten mit manuellen Constraints lösten das Problem unvollständig geschlossener Augen beim Wimpernschlag.

Die Modelle wurden für drei verschiedene Einsatzszenarien entwickelt: Helmkamera, Webcam und Multi-Kamera-Setup. Im Jahr 2022 lag der Fokus auf der Helmkamera, was sich als effektiver Ansatz erwies, da dies durch die fixe Kameraposition relativ zum Gesicht ein vereinfachtes Szenario repräsentiert. Die verschiedenen Probleme und deren Behebung führten zu einer komplexen Trainings-Pipeline, die diverse manuelle Schritte erforderte. Die Entwicklung einer automatischeren Pipeline trug zu einer höheren Entwicklungsgeschwindigkeit bei, obwohl die Komplexität der Pipeline aufgrund der Schwierigkeit, annotierte Aufnahmen zu erzeugen, weiterhin manuelle Schritte erforderte.

AP 4.9 Temporale Konsistenz von Gesichtstracking (4PM):
Um temporale Konsistenz zu erreichen wurden zwei Hauptansätze verfolgt:

- **Frequenzadaptiver Tiefpassfilter:** Ein Echtzeit-Filter wurde implementiert, um Gesichtsbewegungen zu glätten und die temporale Konsistenz zu verbessern. Der Filter ist darauf eingestellt, langsame Bewegungen stärker zu glätten als Schnelle. Dadurch wird die Latenz bei schnellen Bewegungen minimiert und das wahrgenommene Rauschen, das primär bei langsamen oder statischen Gesichtspartien sichtbar ist, minimiert.

- **Wobble-Fix:** Dieser Ansatz nutzte einen realen Datensatz aus sehr diversen, nicht annotierten Posen. Dieser Datensatz wurde mit dem aktuellen Modell annotiert, und anschließend wurden Augmentationen (Verzerrung, Translation, Unschärfe, Rauschen) auf die Bilder angewendet und mit den generierten Ground-Truth-Annotationen weitertrainiert. Dies führte zu einer leichten Reduktion der Fidelity, aber zu einer deutlich verbesserten temporalen Konsistenz.

Insgesamt hat die Captury GmbH die in der ursprünglichen Vorhabenbeschreibung festgelegten Ziele nicht nur erreicht, sondern in einigen Bereichen, insbesondere durch die Entwicklung separater Modelle für unterschiedliche Kameratypen und die gezielte Behebung von Qualitätsproblemen, übertroffen.

2. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die geleisteten Projektarbeiten von The Captury waren für das Erreichen der Gesamtziele des HiAvA-Projekts von entscheidender Bedeutung. Ohne die Entwicklung des responsiven Trackingsystems wäre eine skalierbare und qualitativ hochwertige Verkörperung von Avataren in Social VR nicht möglich gewesen. Die Arbeiten adressierten direkt die Desiderate unzureichender Skalierbarkeit, Reproduktionsgüte und Komfort bestehender Lösungen.

Die Angemessenheit der Projektarbeiten zeigt sich in der innovativen Herangehensweise:

- Die Fokussierung auf **synthetische Trainingsdaten** in AP 4.5 erwies sich als äußerst effizient, um die enorme Datenmenge zu generieren, die für das Training robuster KI-Modelle erforderlich ist, ohne aufwändige manuelle Annotationen echter Aufnahmen in gleichem Umfang zu benötigen.
- Die **iterative Verfeinerung der KI-Modelle (AP 4.8)** durch einen engen Feedback-Loop zwischen Training, Evaluation und gezielter Datenaufnahme (AP 4.4, AP 4.5) war entscheidend, um spezifische Qualitätsprobleme zu identifizieren und zu beheben.
- Die Entwicklung von **zwei spezialisierten Gesichtstracking-Modellen (AP 4.3)** für Helmkameras und statische Kameras, anstatt eines einzigen generischen Modells, war eine pragmatische und effektive Anpassung, um die unterschiedlichen Anforderungen der Hardwareumgebungen optimal zu erfüllen.
- Das **Blendshape Retargeting (AP 4.6)** und die Verbesserung der **temporalen Konsistenz (AP 4.9)** sind direkte Beiträge zur Glaubwürdigkeit und Natürlichkeit der Avatar-Interaktionen, die über die reine Erfassung von Bewegungen hinausgehen und die immersive Erfahrung maßgeblich verbessern.

Der Personaleinsatz, insbesondere im letzten Projektjahr, wurde aufgrund der Erkrankung eines Projektmitarbeiters flexibel angepasst, ohne die erfolgreiche Durchführung des Projekts zu gefährden. Dies unterstreicht die Robustheit der Projektplanung und des Konsortiums.

3. Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses

Die im Rahmen des Projekts von The Captury entwickelten Technologien haben vielfältige Nutzenpotenziale und hervorragende Verwertungsmöglichkeiten:

Wissenschaftlich-technische Verwertbarkeit:

- **Grundlagen für weitere Forschung:** Die entwickelten KI-Modelle, Datensätze und Tracking-Verfahren bilden eine solide Grundlage für weitere Forschungen im Bereich des Human-Computer Interaction, der Computergraphik und der Künstlichen Intelligenz. Insbesondere die Methoden zur Generierung synthetischer Daten und zur robusten Gesichtserkennung unter verschiedenen Bedingungen sind von hohem wissenschaftlichem Wert.
- **Beitrag zur akademischen Gemeinschaft:** Die Projektpartner, insbesondere JMU und TUD, setzen die von Captury entwickelte Technologie auch über das Ende von HiAvA hinaus in wissenschaftlichen Projekten ein und veröffentlichen ihre Arbeiten.

Wirtschaftliche Verwertbarkeit:

- **Erweiterung des Produktportfolios:** Die entwickelten Gesichts- und Handtracking-Lösungen erweitern das bestehende Produktportfolio von The Captury erheblich und ermöglichen den Einstieg in neue Märkte wie Social VR, immersive Telepräsenz, Broadcast TV und medizinische Anwendungen mit hohen Anforderungen an die Mimik- und Gestenerfassung.
- **Wettbewerbsvorteil:** Die einzigartige Kombination aus Echtzeitverarbeitung, niedriger Latenz und der Fähigkeit, sowohl Ganzkörper- als auch detaillierte Gesichts- und Handbewegungen nicht-invasiv zu erfassen, verschafft The Captury einen signifikanten Wettbewerbsvorteil gegenüber bestehenden Lösungen.
- **Lizenzen und Partnerschaften:** Die Technologie kann an andere Unternehmen lizenziert werden, die im Bereich XR-Hardware, Game-Engines oder Content-Erstellung tätig sind. Strategische Partnerschaften mit Herstellern von HMDs oder mobilen Endgeräten sind denkbar, um die Integration zu erleichtern.
- **Anwendung in verschiedenen Industrien:**
    - **Entertainment & VR Konzerte:** Verbesserte Immersion in Social-VR-Spielen, realistische Avatare für virtuelle Konzerte oder Events.
    - **Medizin:** Präzises Tracking von Mimik und Gestik für Diagnostik und Therapie (z.B. in der Schlaganfallrehabilitation oder Psychotherapie).

    - **Bildung & Training:** Realistische Interaktionen in virtuellen Lernumgebungen, z.B. für Kommunikationstrainings oder Simulationen im Rahmen von Trainings der Polizei, der Feuerwehr und militärischer Spezialeinheiten.
    - **Kollaboration:** Hochwertige und komfortable virtuelle Meetings, die physische Distanzen überwinden und soziale Präsenz erhöhen.
- **Standortbezogene Verwertung:** Die Entwicklung und Vermarktung der Technologien stärkt den Innovationsstandort Deutschland und schafft hochqualifizierte Arbeitsplätze in der Region Saarbrücken.

Konkrete Planungen für die nähere Zukunft:

- **Weiterentwicklung:** Es ist nach Ende des Projektes ein neuer Mitarbeiter für die Weiterentwicklung des Gesichtstrackings eingestellt worden. Zusätzlich wird der erkrankte Mitarbeiter in der nahen Zukunft wieder anfangen zu arbeiten und sich den Bemühungen der Weiterentwicklung anschließen.
- **Forschungskooperationen:** Fortsetzung der Zusammenarbeit mit akademischen Partnern zur Weiterentwicklung der Technologien und Erschließung neuer Anwendungsfelder.

4. Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit wurde der Fortschritt im Bereich des Gesichts- und Körpertrackings bei anderen Stellen genau beobachtet. Insbesondere große Technologieunternehmen wie Meta (Oculus), Google und Apple investieren weiterhin massiv in XR-Technologien und entwickeln eigene Trackinglösungen.

- **Apple (ARKit):** Apples ARKit bietet exzellentes Face-Tracking auf mobilen Endgeräten. Die HiAvA-Ergebnisse, insbesondere das personalisierte Blendshape-Modell und das Retargeting-Verfahren, ergänzen ARKit durch die Möglichkeit, diese Daten auf beliebige Avatar-Modelle zu übertragen und eine konsistente Qualität über verschiedene Hardware hinweg zu gewährleisten. Auch ist ARKit auf Kameras direkt vor dem Gesicht spezialisiert. Das Ziel von Captury ist es, hochwertiges Gesichtstracking auch aus entfernten Kameras, wie sie in einem Motion Capture Setup typisch sind, zu realisieren.
- **Forschungsinstitute:** Es gibt weiterhin umfangreiche Forschungsarbeiten zu markerlosem Tracking und der Synthese von Verhaltensmodellen. Die Stärke von HiAvA lag in der Verknüpfung dieser Ansätze zu einem skalierbaren und alltagstauglichen Gesamtsystem.

Die The Captury GmbH ist dank ihrer engen Vernetzung mit relevanten Forschungsgruppen und Unternehmen über die Fortschritte informiert und konnte ihre

Entwicklungsstrategie entsprechend anpassen, um einen technologischen Vorsprung zu sichern.

5. Veröffentlichungen und Vorträge

Im Berichtszeitraum wurden keine externen Veröffentlichungen oder Vorträge direkt aus den Arbeiten von Captury im Rahmen des HiAvA-Projekts durchgeführt. Dies ist auf die strategische Entscheidung zurückzuführen, sich zunächst auf die Entwicklung und Absicherung der Kerntechnologien zu konzentrieren und vertrauliche Forschungsergebnisse zu schützen. Auch legt Captury, als kleines Wirtschaftsunternehmen, den Fokus auf kommerzielle Verwertung.