

## vTrait

visual Tracking of Arbitrary Targets

<b>Programm / Ausschreibung</b>	Bridge, Bridge_NATS, Bridge_NATS 2019	<b>Status</b>	laufend
<b>Projektstart</b>	01.06.2020	<b>Projektende</b>	31.05.2022
<b>Zeitraum</b>	2020 - 2022	<b>Projektlaufzeit</b>	24 Monate
<b>Keywords</b>	Filmkamera; Tracking; Computer Vision; Machine Learning		

### Projektbeschreibung

Die Zielstellung des aktuellen Forschungsvorhabens ist die robuste visuelle Nachverfolgung eines beliebigen benutzerdefinierten Zielobjekts basierend auf RGB-D-Daten. Diese Funktionalität wird in der Kinematographie benötigt, um eine kontinuierliche Nachführung der Fokusebene auf das scharf abzubildende Motiv durchzuführen, wobei eine ausreichende Zuverlässigkeit auch bei Anwesenheit von Störelementen gewährleistet sein muss. Ausgehend von der zunehmenden Kapazität von gelernten tiefverzweigten Netzwerkrepräsentationen beabsichtigen wir, (i) anhand Siamesischer Lernverfahren Objektähnlichkeiten zu lernen, um initiale Segmentierungsmasken zu generieren und diese (ii) mit unserem bereits entwickelten eingebetteten Stereo-Verfahren zu einer genaueren Tiefensegmentierung zu kombinieren. Diese Kombination von ähnlichkeits- und tiefenbasierten Objektmasken ermöglicht eine zuverlässigere Trennung von Vorder- und Hintergrund und stärkt darüber hinaus Robustheit und Genauigkeit bei visuellen Nachverfolgungsaufgaben. Die erwartete Anwendbarkeit dieser Funktionalität geht dabei weit über die Kinematographie hinaus: Objektverfolgung für autonome Systeme und Handhabungsaufgaben im Bereich der Mobil- und Industrierobotik sind weitere Anwendungsbereiche, die von einer robusten RGB-D-Nachverfolgung einzelner bzw. mehrerer beliebiger Ziele profitieren würden.

### Abstract

The task-specific goal of the current proposal is the robust visual tracking of a user-specified arbitrary target in a cinematographic setting from RGB-D data. Such a scheme shall automatically maintain camera focus on the selected motive, even in the presence of distractors. Relying on the recent representational power of learned deep distributed representations, we intend to combine (i) similarity (Siamese)-learning yielding segmentation proposals and (ii) depth data segmentation from our previously elaborated embedded stereo depth computation scheme. This combination of similarity-based and depth-based segmentation proposals shall greatly robustify foreground-background separation and target association quality. The expected impact of proposed functionality goes beyond cinematography: robotic tracking/grasping and object tracking in autonomous driving are further domains where robust RGB-D tracking of single/multiple arbitrary targets are desired.

## **Projektkoordinator**

- AIT Austrian Institute of Technology GmbH

## **Projektpartner**

- qinematiq GmbH