仕　様　書

1. 件名

一人称視点映像に対する部品位置および組立状態のアノテーション作業

2. 研究の概要

国立研究開発法人産業技術総合研究所人工知能研究センター（以下、「産総研」という。）では研究開発成果の社会実装への橋渡しプログラム（BRIDGE）AI×ロボット・サービス分野の実践的グローバル研究の一環として「映像からの部品単位の組立行動理解およびその作業者支援への応用」（以下、「本課題」という。）を担っている。ここでは家具や電子部品等の製品を人が組み立てる・分解する・操作する状況において、映像中に出現する部品の位置およびその組立状態（部品同士の結合関係）を認識する人工知能技術の開発を目指している。この一環として、訓練および評価のためのアノテーション作業を行う必要がある。

3. 作業の概要

特定の種類の物体の状態変化が含まれる動画ファイル 1,461 件について、動画中の部品位置およびその組立状態に関するアノテーションを行うこと。

4. 作業項目

4-1. 動画ファイルの取扱
4-2. アノテーション作業の流れ
4-3. 部品位置および組立状態のアノテーション
4-4. アノテーション結果ファイルのレビュー

5. 作業項目別仕様内容

5-1. 動画ファイルの取扱

動画ファイルはインターネット上に公開されているものであるが、その内容そのものは本課題の実施にあたっての非公開情報であるため、外部に漏洩しないよう、慎重に取り扱う必要がある。そのため、下記の事項を遵守すること。

・受注者側では作業責任者を選任し、受注者および再委託先の事業者に係る実施体制図（各法人名およびその役割を記したもの）を産総研に提出すること。実施体制図に記した事業者の外に画像ファイルを流通させることを厳禁する。
・動画ファイルは、産総研から受注者が指定するクラウドに、暗号化された通信手段（例えばSSL）によって転送するものとする。
・受注者側のクラウドにおいては、先の実施体制図に記載された事業者内の人員以外のアクセスは禁止すること。
・外部からの不正アクセスによって情報漏洩した疑いがある場合には、速やかに産総研にその旨を連絡すること。

5-2. アノテーション作業の流れ

アノテーション作業は、以下の流れで行うこと。

1) 産総研から動画ファイルを受領する。
2) 1 つの動画ファイルについて、部品位置および組立状態のアノテーションを行う。
3) 2)の結果を産総研に納品する。

4) 産総研にて、2)の結果をレビューして、修正点があれば差し戻しをする。

5-3. 部品位置および組立状態のアノテーション

5-3-1. 背景

本研究では、製造業などの分野で作業者が手や道具を用いて製品を組み立てる活動において、作業者の作業を写した映像から (i) 作業者の組立行動 および (ii) 製品の部品単位での組立状態を認識する人工知能技術を開発している。

今回、その題材として多様な製品の組立・分解の様子を収録したデータセットである HoloAssist データセット[Wang+, ICCV'23]を選定した。本データセットには、19 のタスクが収録されているが、今回はそのうち 8 製品、11 タスク、1,461 試行、合計時間 106.6 時間分の映像に対して各部品の位置および組立状態の変化のアノテーションを行いたい。

5-3-2. 用語説明

本項では、作業内容の説明に必要な重要な用語の説明を行う。

「タスク」とは、当該動画中で組立または分解される製品およびその指示の組からなるユニークな作業を指す。

「物体」とは、各製品を構成する本体・部品および製品と相互作用する物体(例:SD カード、ドライバー)を指す。タスク毎にアノテーションすべき物体は指定される。レンチやドライバーといった製品の組立に使用する道具についてはアノテーションの対象外とする。詳細は「別添資料 1(アノテーションポリシー)」を参照のこと。

「領域」とは、画像中のある物体の位置(バウンディングボックスまたはセグメンテーションマスク)を指す。

「関係」とは、2 つ以上の物体の間において、その物体の意図された機能によって規定された幾何的関係のことを指す。今回は、意図しない機能の発生に基づく幾何的関係(汎用の箱の中に適当なものを入れる、椅子の上に物を置く)は関係とみなさない。

「状態」とは、ある 1 つの物体またはその一部の領域がその物体の意図された機能によって規定された幾何的・意味的状態を指す。

5-3-3 物体・関係・状態の種類

表 1 に物体、関係、状態のそれぞれの種類を示す。原則として、表に記載の物体・関係・状態についてのみアノテーション行うこと。ただし、表 2 に未記載の物体のうち、タスクの遂行に関係すると思われる物体、またはタスクの遂行に関係すると思われる関係・状態があると作業者が判断した場合には、その旨を産総研に報告し指示を仰ぐこと。産総研が適当であると認めた場合は、以後それらについてもアノテーションを行うこと。

5-3-4. アノテーションの対象

表 3 にタスク別の動画数および長さ、表 4 に動画ファイルの仕様を示す。下記に示すフレームにアノテーションを付与すること。

- A:対象物体に左右のいずれかの手が初めて接触するフレーム
- B:フレーム A 以後に、ある関係・状態を持つ物体が初めて登場するフレーム
- C:ある物体間の関係または状態が変化したフレーム

ただし、別に定める例外について、アノテーションを不要とする場合がある。詳細については

別添資料 1（アノテーションポリシー）を参照のこと。
数量については実際の動画内容に応じて変化する場合があるが、表 2 左段に示す目安対象フレーム数および動画数から算出した合計想定フレーム数は 43,989 フレームである。全動画または対象フレーム数が同数量に達するまでアノテーションを行うこと。

5-3-5. 物体種類・位置のアノテーション
5-3-4 の条件を満たすフレームについて、5-3-3 に示す物体カテゴリの種類および位置（セグメンテーションマスク）を付与すること。1 動画内において、同一のインスタンスには同一の ID を付与すること。

5-3-6. 関係・状態のアノテーション
5-3-5 で付与した各物体のセグメンテーションマスクまたはその組について、5-3-3 に示す関係・状態について、当該フレームで変化のあった関係および状態を付与すること。ただし、A のフレームについては当該フレームに出現する全物体のすべての関係および状態、B のフレームについては当該物体のすべての関係および状態について値を付与すること。

5-4. アノテーション結果ファイルのレビュー
5-3-5、5-3-6 のアノテーションについて産総研がレビューを行うこととする。アノテーション結果に修正点が生じれば、次のレビューまでに修正を行うこと。最終的な納品時点では、すべてのアノテーション結果についてレビューが済んでいることとする。

6. 納品の完了
「「8.納入物品」に記載された納入物品が過不足なく納入され、仕様を満たしていることの確認をもって、納入の完了とする。」

7. 特記事項
(1) 産総研は、受注者がアノテーションにより得られた技術上の成果のうち産総研が指示するもの（以下「成果」という。）についての利用及び処分に関する権利を専有するものとする。
(2) 受注者は、成果に係るアノテーションの著作権を産総研に無償で譲渡するものとし、著作者人格権を行使しないものとする。
(3) 受注者は、検収終了後、直ちに別紙様式（本文書末尾に記載）による著作者財産権譲渡証書及び著作者人格権不行使証書を提出するものとする。
(4) 受注者は、産総研に対し、納品した成果品が第三者の著作権を侵害しないことを保証するものとする。なお、納品した成果品について、第三者の権利侵害の問題が生じ、その結果、産総研又は第三者に費用や損害が生じた場合は、受注者は、その責任と負担においてこれを処理するものとする。

8. 納入物品
以下の納品物を電子ファイルとして、ファイル転送サービスによって納品すること。
1）json 形式での各動画へのアノテーション情報 1,461 件

9. 納入期限及び納入場所
納入期限：2025 年3月31日
納入場所：〒305-8560 茨城県つくば市梅園 1-1-1
つくば本部・情報技術共同研究棟 4 階 4301 室

10. 付帯事項

- 本仕様書の技術的内容及び知り得た情報については、守秘義務を負うものとする。
- 本仕様書の技術的内容に関しては、請求担当者の指示に従うこと。また、本仕様書に定めのない事項及び疑義が生じた場合は、調達担当者と協議のうえ決定する。

参考文献

[Wang+, ICCV'23] Wang, Xin, et al. "Holoassist: an egocentric human interaction dataset for interactive ai assistants in the real world." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

表 1 関係・状態の種類およびその典型的な値

※値は一例であり、実際に取りうる値についてはタスク・物体毎に定める。判断が困難なケースについては個別に産総研の判断を仰ぐこと。

名称	値
Connection（関係）	Attached：物体 A が物体 B と確実に結合している Contact：物体 A が物体 B と接触している、結合の中途である、確実に支持されていない（例：力を加えた場合にずれる） ※ 物体同士が重なっているだけといった、その物体の意図する結合関係と無関係の接触は separated とすること。例えば、ねじ穴を通して部品同士が組み合わさる場合、そのねじ穴周辺以外での接触は対象外とする。 ※ ただし、部品の向きの間違いなど、正しい結合関係でない場合には正しい場合と同様 contact/attached ラベルを付与すること。 Separated：物体 A が物体 B と触れていない、結合位置と別の場所で接触している
Containment（関係）	Filled：物体 B（一般に液体）が物体 A（一般に容器）に入っている（量の多寡を問わない） Empty: 物体 B（一般に液体）が物体 A（一般に容器）に入っていない
Access（状態）	Open (active)：ある物体のふたなどが空いている、機能が有効になっている Close (inactive)：ある物体のふたなどが閉じている、機能が無効である
Inclusion（状態）	Inserted：ある物体の内部に別の物体が挿入されている Empty：ある物体の内部に意図された物体が挿入されていない、空である、挿入または抜去の途中である
Switch（状態）	On：ある物体の機能が有効になっている、ボタンが押されている Off：ある物体の機能が無効である、切り替えの中途である、ボタ

	ンが押されていない
Power（状態）	On：ある物体の電源が入っている Off：ある物体の電源が入っていない ※物理スイッチが見える場合 Power 相当のものが Switch 扱いになる場合がある

表 2 タスク毎の物体・関係・状態の一覧

タスク	物体	関係（物体 1, 物体 2）、または状態（物体）およびその値
GoPro（アクションカメラ） 対象フレーム数目安 26（6 分 30 秒）	camera battery camera mount head mount sd card buckle mount selfie stick tripod screw power button	Connection ● camera-[camera/head/buckle] mount ● camera-selfie stick ● camera-tripod ● camera-screw ● [camera/head/buckle] mount-screw ● selfie stick-screw ● tripod-screw Access ● camera, lid_open/lid_close ● camera, finger_open/finger_close Inclusion ● camera, battery_inserted/battery_empty Power ● camera, power_on/power_off
Switch（ゲームコントローラー） 対象フレーム数目安 27（2 分 57 秒）	console blue controller red controller joy-con grip game card micro sd card	Connection ● console-blue controller ● console-red controller ● joy-con grip-blue controller ● joy-con grip-red controller Access ● console, memory_lid_open/memory_lid_closed ● console, stsand_active/stand_inactive ※ゲームカードの蓋を兼ねる Inclusion ● console, memory_card_insterted/memory_card_empty ● console, game_card_inserted/game_card_empty Power ● console, power_on/power_off

DSLR(一眼レフカメラ) 対象フレーム数目安 17(1 分 51 秒)	Camera Lens Body cap Rear lens cap Front lens cap Battery SD card	Connection ● camera-body cap ● camera-lens ● lens-rear lens cap ● lens-front lens cap Access ● camera, battery_lid_open/battery_lid_closed ● camera, card_lid_open/card_lid_closes Inclusion ● camera, battery_inserted/battery_empty ● camera, card_inserted/card_empty Switch ● camera, switch_on, switch_off
Nespresso(コーヒーメーカー) 17(2 分 2 秒)	Coffee maker Support Capsule tray Plastic cup Paper cup Mug Coffee capsule Water Coffee	Connection ● coffee maker-support ● support-capsule tray Containment ● coffee maker-water ● plastic cup-water ● paper cup-water ● mug-water ● plastic cup-coffee ● paper cup-coffee ● mug-coffee Access ● coffee maker, water_lid_open/water_lid_closed ● coffee maker, support_open/support_closed ● coffee maker, handle_up/handle_down Inclusion ● coffee maker, capsule_inserted/capsule_empty ● capsule tray, capsule_inserted/capsule_empty Switch ● coffee maker, button_pressed/button_not_pressed
SmallPrinter(プリンター) 対象フレーム数目安 17(2 分 18 秒)	Printer Paper ink	Connection ● printer-ink Access ● printer, drawer_open/drawer_close ● printer, output_tray_open/output_tray_close ● printer, scsanner_tray_open/scanner_tray_close ● printer, scanner_input_tray_set/scanner_input_tray_unset ● printer, scanner_output_tray_set/scanner_output_tray_u

		nset ● printer, input_tray_open, input_tray_closed ● printer, input_tray_set/input_tray_unset ● printer, output_tray_set/output_tray_unset ● printer, front_panel_open, front_panel_closed Switch ● printer, front_button_on/front_button_off ※暫定で正面パネルのボタン/ディスプレイのいずれかの選択 Power ● printer, power_on/power_off
Rashult_assemble (ワゴンの組立て) 対象フレーム数目安 83(22 分 17 秒)	roller (4x) bolt (6x) nut (6x) screw (8x) straight pipes (4x) bended pipe (4x) basket (3x) cover (6x)	Connection ● straight pipes-straight pipes ● straight pipes-screw ● bended pipe-bended pipe ● straight pipes-bended pipe ● bended pipe-screw ● bolt-cover ● bolt-straight pipes ● bolt-nut ● straight pipes-basket ● basket-nut ● basket-basket ● roller-bended pipe
Rashult_disassemble(ワゴンの分解) 対象フレーム数目安 同上	同上	同上
Gladom_assemble (トレーテーブルの組立て) 対象フレーム数目安 15(3 分 51 秒)	Long screw Short screw (3x) Legs (2x) Support ring Tray	Connection ● long screw-legs ● legs-legs ● short screw-legs ● short screw-support ring ● top plate-support ring
Gladom_disassemble(トレーテーブルの分解) 対象フレーム数目安 同上	同上	同上

Marius_assemble (椅子の組立て) 対象フレーム数目安 33(6 分 5 秒)	Screw (4x) Connector (4x) Bolt Nut Leg (2x) Top plate	Connection ● bolt-leg ● bolt-nut ● leg-leg ● nut-leg ● screw-connector ● screw-leg ● screw-plate ● connector-leg ● leg-plate

表 3 タスク毎の動画数および合計長(分)

タスク	動画数	合計長(分)
GoPro(アクションカメラ)	162	805
Switch(ゲームコントローラー)	159	684
DSLR(一眼レフカメラ)	164	417
Nespresso(コーヒーメーカー)	120	318
SmallPrinter(プリンター)	113	430
Rashult_assemble(ワゴンの組立て)	84	1359
Rashult_disassemble(ワゴンの分解)	104	776
Gladom_assemble(トレーテーブルの組立て)	135	460
Gladom_disassemble(トレーテーブルの分解)	153	305
Marius_assemble(椅子の組立て)	123	534

Marius_disassemble（椅子の分解）	144	308
合計	1,461 試行	6396 分（106.6 時間）

表 4 動画ファイルの仕様

属性	値
解像度 [pixel]	896×504
フレームレート[FPS]	通常 30（撮影条件により変動）

別添資料 1 アノテーションポリシー

■ 動画データ（HoloAssist）の概要

本動画群は、多様な製品の組立・分解の様子を収録したデータセットである HoloAssist データセット[Wang+, ICCV'23]の一部のタスクより抜粋したものである。各動画は AR ゴーグルである Microsoft HoloLens を用いて撮影されており、装着者の視点から見た作業映像が記録されている。
各タスクでは、作業者は別の指示者の指示を受けながら、表 2 に示す製品の操作・組立・分解のいずれかを行う。例えば GoPro タスクでは、「バッテリーを取り返す」「マイクロ SD カードを取り換える」「電源を入れる/落とす」「自撮り棒または三脚に取り付ける」等の行動が含まれる。作業の手順は指示者に委ねられており、その内容は試行毎に若干異なる場合がある。作業者がタスクに不慣れな場合、その意図した手順通りに実行されず、誤った行動を行う場合がある。

■ アノテーション対象フレームの選択

下記に示すフレームにアノテーションを付与すること。
- A:対象物体に左右のいずれかの手が初めて接触するフレーム
- B:フレーム A 以後に、ある関係・状態を持つ物体が初めて登場するフレーム
- C:ある物体間の関係または状態が変化したフレーム

注意事項:
- 関係・状態の定義については表 1 を参照のこと。
- ただし、以下の場合についてアノテーション対象から除外する:
  - ある関係または状態が短時間（概ね 1 秒以内）の間に再度元の状態に戻る場合、
  - 上記について、1 秒以上別関係・状態にあった場合でも、その物体への関与が連続的に行われていると判断できる場合
- A-C の各ケースについて、該当する最初のフレームの (i) 画面全体または物体が著しくぶれている (ii) その関係・状態の変化に関与する物体の視認が困難である場合、その直後（0.5 秒以内）で(i)(ii)の程度が少ない適当なフレームを選択すること。
- 同様に、A-C の各ケースについて、関係または状態の変化が画面外で発生したことが動

画より明らかになった場合、その事実が判明した直後で当該物体が大きなぶれや欠けがなく映っているフレームを選択すること。

■ 物体種類および位置のアノテーション

前項の条件を満たすフレームについて、表 2 に示す物体カテゴリの種類および位置（セグメンテーションマスク）を付与すること。1 動画内において、可能な範囲で同一のインスタンスには同一の ID を付与すること。

注意事項:

- セグメンテーションについては、1 ピクセル単位での正確性を求めるものではなく、各物体のおおよその範囲が示されていればよい。
- 手によるオクルージョンなどに伴い、1 つの物体のセグメンテーションが 2 つ以上の領域に分割される場合には、それらをひとまとめとして 1 つの物体とカウントする。
- 手または他の物体によって一部の領域が隠れる場合、原則として見える箇所にのみ領域アノテーションを付与する。ただし、関係または状態の変化を伴うフレームにおいて、当該物体が見えていないが明らかに関係または状態の変化が起こっている場合に限り、その物体があると思われる領域をマークすること。
- 表 2 に今回出現する物体を網羅しているが、試行によっては表中に存在しないが、組立状態にかかわる物体が登場する場合がある。その場合は、その旨を産総研に報告し指示を仰ぐこと。産総研が適当であると認めた場合は、以後それらについてもアノテーションを行うこと。
- 各物体の同一性について、当該物体が画面外に出るなどして判定が困難な場合には新規に ID を振りなおしてもかまわない。

■ 関係・状態のアノテーション

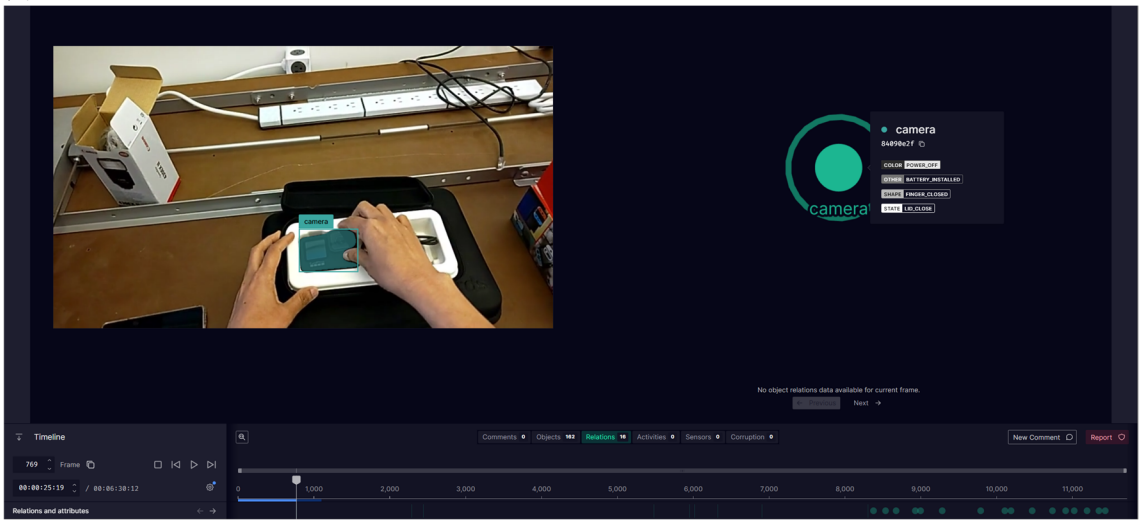
前項で付与した各物体のセグメンテーションマスクまたはその組について、表 2 に示す関係・状態について、**当該フレームで変化のあった関係または/および状態**を付与すること。ただし、A のフレームについては**当該フレームに出現する全物体のすべての関係および状態**、B のフレームについては**当該物体のすべての関係および状態**について値を付与すること。

注意事項:

- 表 2 に今回登場すると思われるすべての関係・状態を記述しているが、データセットの性質上網羅しきれていない場合がある。新規の関係・状態ラベルを与える必要が場合には、表 1 の定義に従ってもっとも適当なものを選ぶこと。
- 関係・状態のアノテーション基準に不明点がある場合にはその旨を産総研に申し出、その指示に従うこと。
- Connection・containment・insertion などの関係・状態について、必ずしも映像中の手がかりからは確定できない場合（例：ねじが規定位置まで確実に締まったがどうか）がある。そうした場合には、作業者の前後の文脈を含めて総合的に確実に結合した（例：増し締めが以降行われていない、ケースがスムーズに閉まった）と判断したタイミングで付与すること。

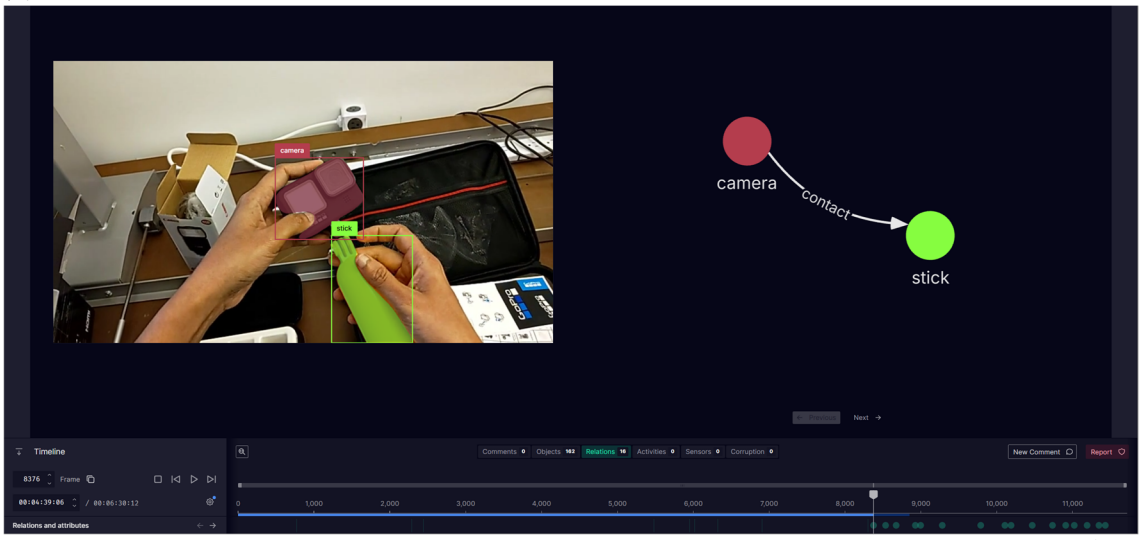
■ 具体例

例 1:



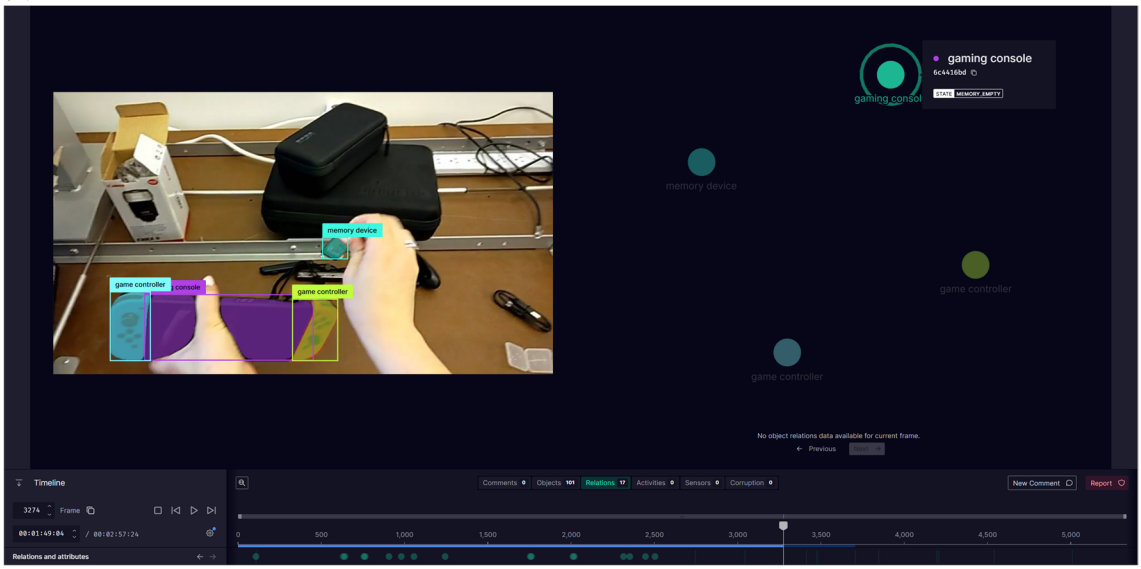
動画中で初めてその物体に触れる瞬間(ケース A)である。ここで映っている物体はカメラのみである。最初の出現フレームであるため、アノテーション対象のすべての状態の初期値が付与されている。

例 2:



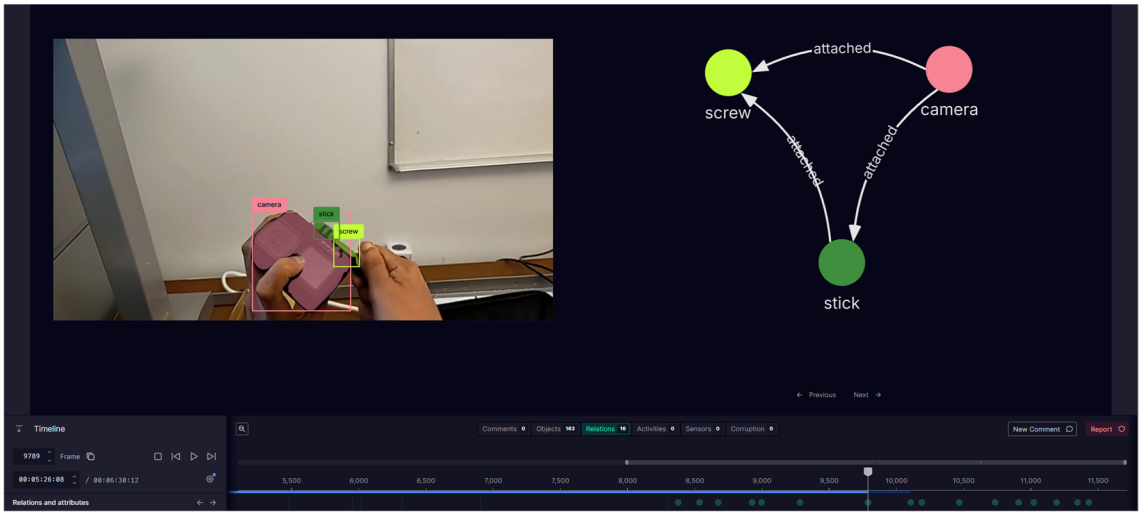
関係(Connection)の変化を伴うフレーム(ケース C)である。ここでは、カメラ(camera)と自撮り棒(selfie stick)が separated→contact になった瞬間であるためその変化分のみ付与する。Camera と selfie stick の位置はこの時点では揃っていないが規定の位置への装着を試みていることからこのフレームに付与した。

例3:



状態(containment)の変化を伴うフレーム(ケース C)である。Console の状態が inserted→empty に変化したためそのラベルを付与している。実際にメモリーカード(sd card)が取り除かれたのは数フレーム前であったが、物体がぶれてはっきり映っていなかったため、sd card が明瞭になる瞬間を対象とした。

例 4:



関係(Connection)の変化を伴うフレーム(ケース C)である。ここでは。カメラ(camera)と自撮り棒(selfie stick)がねじ(screw)を介して確実に結合された(attached)瞬間である。こうした場合は互いに接触している 2 者関係全てについて attached とする。Screw の先は隠れているため、セグメンテーション上は持ち手の部分しかセグメンテーションを付与していない。

別紙様式

年　　月　　日

著 作 者 財 産 権 譲 渡 証 書

国立研究開発法人産業技術総合研究所　殿

受　注　者
住　　所
会　社　名
代表者氏名　　印

アノテーション受注契約　　（　　年　　月　　日　契約）
件　名

上記契約により作成したアノテーションの所有権及び著作権（著作権法第 27 条及び第 28 条に規定する権利を含む）は、国立研究開発法人産業技術総合研究所に譲渡したことに相違ありません。ただし、自己所有していた権利は除くものとします。

【注：契約書を取り交わす場合、著作者財産権譲渡証書が重複することになるため、仕様書の添付様式を取り除くこと。】

別紙様式

年　　月　　日

著作者人格権不行使証書

国立研究開発法人産業技術総合研究所　殿

受　注　者
住　　所
会　社　名
代表者氏名　　印

アノテーション受注契約　　（　　年　　月　　日　契約）
件　名

上記契約により作成したアノテーションの著作権（著作権法第27条及び第28条に規定する権利を含む）に係わる著作者人格権を行使しないことを約束します。

なお、著作者人格権を行使しようとする場合は、国立研究開発法人産業技術総合研究所の承認を得るものとします。

【注：契約書を取り交わす場合、著作者人格権行使証書が重複することになるため、仕様書の添付様式を取り除くこと。】