Body-worn Camera Object Detection Annotation Instructions

1. Summary

For this use case objects appearing in a set of body-worn camera images need annotation.

1.1. Annotate the following targets across all frames:

There are only minimal size requirements.

- HUMAN
    - HUMAN_BODY (min. 30 pixels diagonal)
        - **If any portion of a person is visible, regardless of percent visibility, annotate with a bounding box according to the minimum size.**
        - It is possible that there will just be the foot or arm of a person visible and we want that to be annotated.
        - Should you see two arms/hands of a person, such as the person wearing the camera, both arms should be included in one bounding box.
        - **Select the Tag and the AgeType that best describe the human:**
            - **Tag: POLICE (in uniform), CIVILIAN, Unclear.**
            - **AgeType: Child, Teenager, Adult, Senior, Unclear, [NotVisible].**
    - HUMAN_FACE (min. 20 pixels wide)
        - If the face is less than 20 pixels wide, you do not need to annotate it.
        - If the face has no mask/covering, make sure that at least one eye (also includes glasses and sunglasses), half the nose, and half the mouth is visible.
        - If the face has a mask/covering, make sure there is at least one facial feature visible (e.g. mouth, eye, or nose). The eye may be covered by glasses, sunglasses or the mask. The mask/covering can be covering the mouth, nose, or eye. Provide the best estimate of the covered face features including the eyes, nose, mouth, chin, and ears when annotating.
- VEHICLE (min. 30 pixels diagonal)
- LICENSE_PLATE (min. 20 pixels wide)
- SCREEN (min. 30 pixels diagonal)
    - CELL_PHONE
    - MONITOR (includes television, tablets, laptops, in vehicle GPS/radio LCD, and

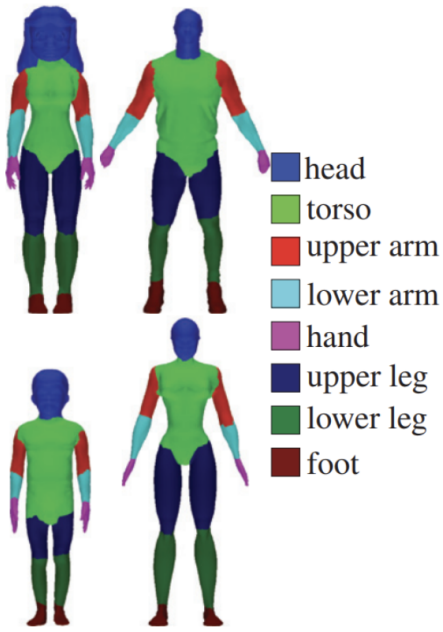
screens used for advertisements). **Bounding boxes should be added around all screens and should include the edges/frames.**
    - **No bounding box should include non-digital keyboards such as computer keyboards.**
    - **Select the appropriate value (TRUE/FALSE) for the attribute "IsIlluminated", which describes if the screen is on or off.**
- WEAPON (min. 20 pixels diagonal).
    - LONG_GUN (includes rifle, machine gun, etc) and SHORT_GUN (includes taser, handgun, pistol, uzi, etc.) are the possible SubClassifications.
        - if a target is unclear or less than 15x15 pixels, do not mark it.
        - Until an object becomes at least 50% visible, do not draw the box.
    - For image annotation we make sure the gun barrel is visible when the gun is held. If the barrel is not visible, do not classify it as a gun. However, guns in holsters are real guns even if their barrel is not visible.
    - When the gun is being held, make sure the grip of the gun occluded by the hand is included inside the box.

1.2. Attributes:

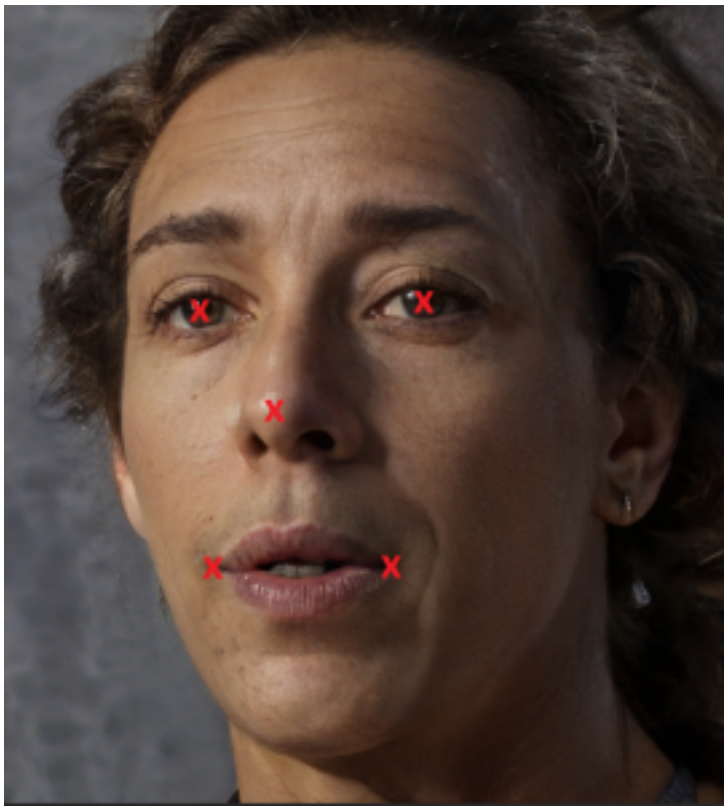
- **PlateText:** License plate text is only applicable to the LICENSE_PLATE class. ***(Use the "Point Label" number 2 to add the license plate text).*** The characters on the left of any specialty plates should be ignored and there should be no whitespace; only characters and digits. In the following example the annotation would be SLPAN and CUREDA and all other characters are ignored.

- **PlateReadable:** TRUE or FALSE indicating if the characters on the license plate are identifiable. When set to TRUE, it is expected that *PlateText* is populated. If for any reason the annotator is guessing more than two characters in total, the PlateReadable attribute should be set to FALSE. If there are two or less characters the annotator is unsure of because they can't decipher similar characters such as D and O, or I and 1, then it is a close guess and the PlateReadable attribute should be labeled as TRUE.
- **Confidence**: boolean, indicating definite clear and accurate plate text (TRUE) vs. less than 100% certain on plate text (FALSE).
- **Occlusion:** A qualitative description of the visibility of the object : NONE, PARTIAL, FULL.
    - NONE: less than 15% of the object is occluded.
    - PARTIAL: between 15% and 85% of the object is occluded.
    - FULL: more than 85% of the object is occluded.
- **HumanCautionTag:** An identifier that specifies internally whether the annotation requires any special attention for redaction.
    - NUDE: One or more intimate parts (buttocks, genitals, and female breasts) are not kept covered by clothing. Undergarments covering the person and cleavage would not be considered nude.
    - NORMAL: All the intimate parts are kept covered by clothing and there is no blood.
    - BLOOD: There is blood visible on the person or on their face.

- **HumanBodyPartVisibility:** Indicates the visible and invisible human body parts and only applies to the HUMAN_BODY subclass. **At least 50% visibility is required for a body part to be deemed visible**. **Invisibility is due to pose, camera view or occlusion by another object**, hence ***parts that are covered by clothing are considered visible***.
  When annotating you should select all the body parts that are visible.
  The body parts are as follows (see the illustration down below):
  - Head
  - Torso
  - LeftUpperArm
  - RightUpperArm
  - LeftLowerArm
  - RightLowerArm
  - LeftHand
  - RightHand
  - LeftUpperLeg
  - RightUpperLeg
  - LeftLowerLeg
  - RightLowerLeg
  - LeftFoot
  - RightFoot

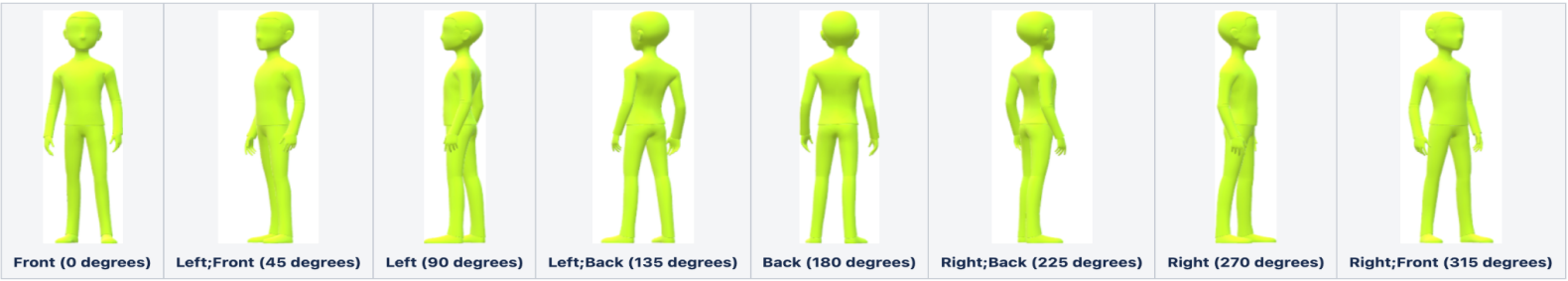


- **FaceKeypointsVisibility:** Indicates the visible and invisible facial key points and only applies to the HUMAN_FACE subclass. **Invisibility is due to pose, camera view or occlusion by another object as well as due to a mask or sunglasses. When annotating, you should select all the facial key points that are visible.**
  The facial key points are as follows (see the illustration down below):
    - LeftEyeIsVisible
    - RightEyeIsVisible
    - NoseIsVisible
    - LeftMouthIsVisible
    - RightMouthIsVisible



- **FaceQuality:** An identifier that classifies the quality of the face into one of the following categories:
    - 1: Poor quality face, poor image quality, or significant pose, or partial occlusion. Even though the face is in the view of the camera, the face is not quite visible and is blurry either due to low resolution or motion blur.
    - 2: Medium quality face. The face is clearly visible, but the face is not frontal facing (the face is not directly facing the camera). Both eyes, the nose, and the mouth have to be visible.
    - 3: Good quality face. Good image quality, frontal, no occlusion. The face is clear and pointing straight at the camera. Two Eyes are clearly visible and straight to the camera.

- **Pose:** An identifier that classifies the pose of the object into any of the canonical poses, e.g., LEFT;BACK.
  There are eight canonical poses describing the view of the object which are demonstrated below. Pose is annotated for both vehicles and human bodies. Note that the Front for a vehicle means the hood while back means the trunk.



- **Posture:** An identifier that classifies the posture of the HUMAN_BODY into either UPRIGHT, SITTING, LYING, or UNCLEAR.

1.3. Attributes per Object type:

For each object type, here is the list of attributes to label

1. HUMAN_BODY: Occlusion, Pose, Posture, HumanBodyPartVisibility, HumanCautionTag, Tag, AgeType.
2. HUMAN_FACE: FaceQuality, FaceKeypointsVisibility.
3. VEHICLE: Occlusion, Pose.
4. LICENSE_PLATE: Occlusion, PlateReadable, Confidence, PlateText.
5. SCREEN: Occlusion, IsIlluminated.
6. WEAPON: Occlusion.