TeamAware

TEAM AWARENESS ENHANCED WITH ARTIFICIAL INTELLIGENCE AND AUGMENTED REALITY

Deliverable D3.1

Dataset report

Editor(s):	Thales SIX GTS France: Andreina Chietera ; TREELOGIC : Victor Medina, Victor Fernandez Carbajales; EUCENTRE : Ilaria Senaldi, Chiara Casarotti
Responsible Partner:	Thales SIX GTS France
Status-Version:	Final v1.0
Date:	03/05/2022
Distribution level (CO, PU):	PU

Project Number:	GA 101019808
Project Title:	TeamAware

Title of Deliverable:	Dataset report
Due Date of Delivery to the EC:	30/04/2022

Workpackage responsible for the Deliverable:	WP3 – Visual Scene Analysis System
Editor(s):	THALES SIX GTS FRANCE SAS
Contributor(s):	TREELOGIC, EUCENTRE
Reviewer(s):	HAVELSAN, EUCENTRE
Approved by:	SIMAVI
Recommended/mandatory readers:	WP04, WP09, WP10, WP11, WP12, WP13

Abstract:	The aim of this task is to gather realistic datasets to train and test AI-based algorithms. In order to provide automatic analysis of the features of the visual scene acquired by the E/O systems, AI-based technologies will rely on an offline training process requiring large amounts of labelled representative data. For each target corresponding to the use cases and requirements, there will be a data acquisitions within defined scenarios and contexts. After the acquisition phase, data will be annotated with the ground truth defined by end-users.
Keyword List:	Victims detection, images segmentation, smoke & fire detection, structural damages.
Licensing information:	The document itself is delivered as a description for the European Commission about the released software, so it is not public.
Disclaimer	This deliverable reflects only the author's views and the Commission is not responsible for any use that may be made of the information contained therein

Document Description

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	13/02/2022	TOC	THALES
v0.2	15/02/2022	Preliminary content	THALES
V0.3	25/03/2022	Drafted contributions to chapter 2, 3 & 4. Content of section 5.	THALES, TREE, EUCENTRE
V0.4	28/03/2022	Refined contributions to chapter 2, 3, 4 & 5.	THALES, TREE, EUCENTRE
V0.5	01/04/2022	Content for the introduction and conclusion	THALES
V0.6	07/04/2022	Minor corrections	TREE, EUCENTRE
V0.7	08/04/2022	References for section 3	THALES
V1.0	03/05/2022	Final version to be released to the EC	THALES

Table of Contents

Document Description 3
Table of Contents 4
List of Figures 6
List of Tables 6
Terms and abbreviations 7
Executive Summary 8
1 Introduction 9
1.1 About This Deliverable 10
1.2 Document Structure 10
1.3 Relation with Other Tasks and deliverables 10
2 Survey on Rescue Victim Datasets for Computer Vision applications 12
2.1 Relevance to Project Scenarios 12
2.2 MPII Human Pose 12
2.3 UAV-Human 13
2.4 OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors 14
2.4.1 Kinect dataset description 14
2.5 HMDB: a large human motion database 15
2.6 Hollywood 3D: Recognising Actions in 3D Natural Scenes 16
2.7 MSRDailyActivity3D 17
2.8 Summary of Dataset for Victim Detection 18
3 Survey on Images Segmentation Datasets for VSAS 19
3.1 Switzerland Aerial Drone Footage Datasets 19
3.1.1 Relevance to Project Scenarios 19
3.2 USTC Forest Smoke and Fire Dataset 20
3.2.1 Relevance to Project Scenarios 21
3.3 The Stanford Drone Dataset 24
3.3.1 Relevance to Project Scenarios 25
3.4 Aerial Dataset of Floating Objects (AFO) Dataset 26
3.4.1 Relevance to Project Scenarios 26
3.5 Low Altitude Disaster Imagery (LADI) Dataset 27
3.5.1 Relevance to project scenarios 29
3.6 Indoor Semantic Dataset (2D-3D segmentation) 29
3.6.1 Relevance to project scenarios 29

3.7 Amazon Web Service Open Data Registry.. 31
3.8 Summary of Image Segmentation Datasets for VSAS 31
4 Survey on Building Defect Datasets 32
4.1 Relevance to Project Scenarios 32
4.2 EUCENTRE’s Dataset of Post-event Damages and Structural Inspections 32
4.3 Report of the Synthetic Dataset for Building Defect Recognition by EUCENTRE................. 34
4.4 Summary of Building Defect Datasets.. 34
5 Requirements for the on the In-site Recording Session.. 35
5.1 Structure of Data Collection Process.. 35
5.1.1 Motivation 35
5.1.2 Composition 35
5.1.3 Collection Process... 36
5.2 Mapping of Data Collection VS Scenario Functionalities 36
6 Conclusions.. 37
7 References... 38

List of Figures

Figure 1. End users' operation in demonstration scenarios. The Red boxes present the main task the dataset of this report has to cover 9
Figure 2 : TeamAware methodology and steps to reach the scenario objectives 11
Figure 3 : Example of the inactivity category in the MPII dataset that could be used for victim detections. Different categories are included in the dataset. In the framework of TeamAware project only few categories as laydown, sitting etc. will be considered for victim detection purpose 13
Figure 4 : Example of multisource data (multi modalities) using an UAV recorded in the UAV-Human Dataset 14
Figure 5 : Sample of data from HMBD: fall floor and sit categories could be used in the TeamAware project 15
Figure 6 : Sample of data from Hollywood3D with the correspondent depth maps 16
Figure 7 : Data sample with RGB, with related depth maps and skeleton sequences 17
Figure 8: Switzerland Aerial Drone footage dataset thumbnails 20
Figure 9: USTC Forest smoke and fire dataset. Smoke training set preview 21
Figure 10: USTC Forest smoke and fire dataset. Fire training set preview 22
Figure 11: USTC Forest smoke and Fire dataset. “Big” test set preview 23
Figure 12: USTC Forest smoke and Fire dataset. “Small” test set preview 24
Figure 13: Stanford drone dataset preview 25
Figure 14: Aerial dataset of floating objects. Set 1 preview 26
Figure 15: Aerial dataset of floating objects. Set 2 preview 27
Figure 16: Aerial dataset of floating objects. Set 3 preview 27
Figure 17: Sample images from the LADI dataset 29
Figure 18: Sample images from the 2D modalities catalogue of the dataset. 30
Figure 19: Dataset in 3D modalities includes: the textured 3D mesh models, semantic annotation and their point clouds 30
Figure 20: Example of an annotated image belonging to EUCENTRE’S dataset (L’Aquila 2009 event, shear damage on unreinforced masonry piers and spandrels) 33

List of Tables

Table 1: Number of test and train sequences for each category in the Hollywood3D Dataset. 17
Table 2: Hierarchical labels use in the annotation of LADI dataset 28

Terms and abbreviations

AFO	Aerial Dataset of Floating Objects
AWS	Amazon Web Service
AI	Artificial Intelligence
CCTV	Closed Circuit Television
CNN	Convolutional Neural Network
EC	European Commission
DL	Deep Learning
FR	First Responder
HAR	Human Activity Recognition
HMDB	Human Motion Database
IMS	Infrastructure Monitoring System
IR	Infrared
LADI	Low Altitude Disaster Imagery
LEA	Law Enforcement Agency
ML	Machine Learning
MPII	Max-Planck Institute for Informatics
PWR	Passive Wi-Fi Radar
RF	Radio Frequency
RGB	Red Green Blue
RGB-D	Red Green Blue with Depth Sensor
SDR	Software Defined Radio
UAV	Unmanned Aerial Vehicle
USTC	University of Science and Technology of China
UWB	Ultra Wide Band
VSAS	Visual Scene Analysis System
WP	Work Package

Executive Summary

As, Computer vision or more in general Deep learning may require larger amounts of training data to perform well data management issues including how to acquire large datasets and how to improve the quality of large amounts of existing data become more and more relevant [1]. If the data used in artificial intelligence (AI) training is not sufficiently diverse, well balanced, appropriate to the context and unbiased, problems such as artificial “AI bias” may arise.

Accurate data collection techniques in the era of Big data gives motivation to conduct as a first step a comprehensive survey of the data collection literature on different tasks appropriate to the TeamAware project that will be merged with data acquired in the specific test bed proposed in TeamAware project.

The main reason that leaded to propose this activity is because the right data contribute to generate the appropriate approach to guarantee the quality of the results. Indeed, good processing is about using the right data and algorithms at the right time.

The first version of the dataset report has the aim to introduce relevant open source datasets useful to train model adapted to manmade and natural disaster scenarios. In order to cover the end-user operational requirement 3 main applications are expected to be covered by the Visual Scene Analysis System (VSAS) system:

1. Victim detection
2. Situational awareness using visual segmentation
3. Damage assessment

According to these 3 main tasks, D3.1 proposes a survey on exploitable dataset in the context of the TeamAware project and a perspective of data collection based on synthetic generation and on collection of data in the test sites that will be developed in a second version of the deliverable.

The organisation of the deliverable is as follows:

- Section 1: context of this deliverable into the overall project.
- Section 2: presents different datasets for victims detections
- Section 3: illustrates a collection of datasets for visual segmentations
- Section 4: presents a mixed datasets of real data and generated images for damage detections of infrastructure
- Section 5: is devoted to the base for the future real data collection

1 Introduction

This document presents a list of possible public datasets that could be used in the training phase and in the evaluation process of some or all of the computer vision algorithms within the TeamAware project. The list of datasets presented here is not intended as an exhaustive and complete repository, but rather as an initial list that can be used to begin work in training detectors for selected elements of the demonstration of the scenarios in Figure 1 described for the project in WP2 and later will be improved in WP13. The scenarios are explained in D2.4 in detail. These scenarios to be covered are concerned with

1. The natural disaster event such as an earthquake, involving damaged buildings, human victims, gas leakage, and fire and smoke in an underground station.
2. Human-made scenario consisting mainly in a terrorist attack with further incidents (e.g. explosions, toxic chemical attacks).

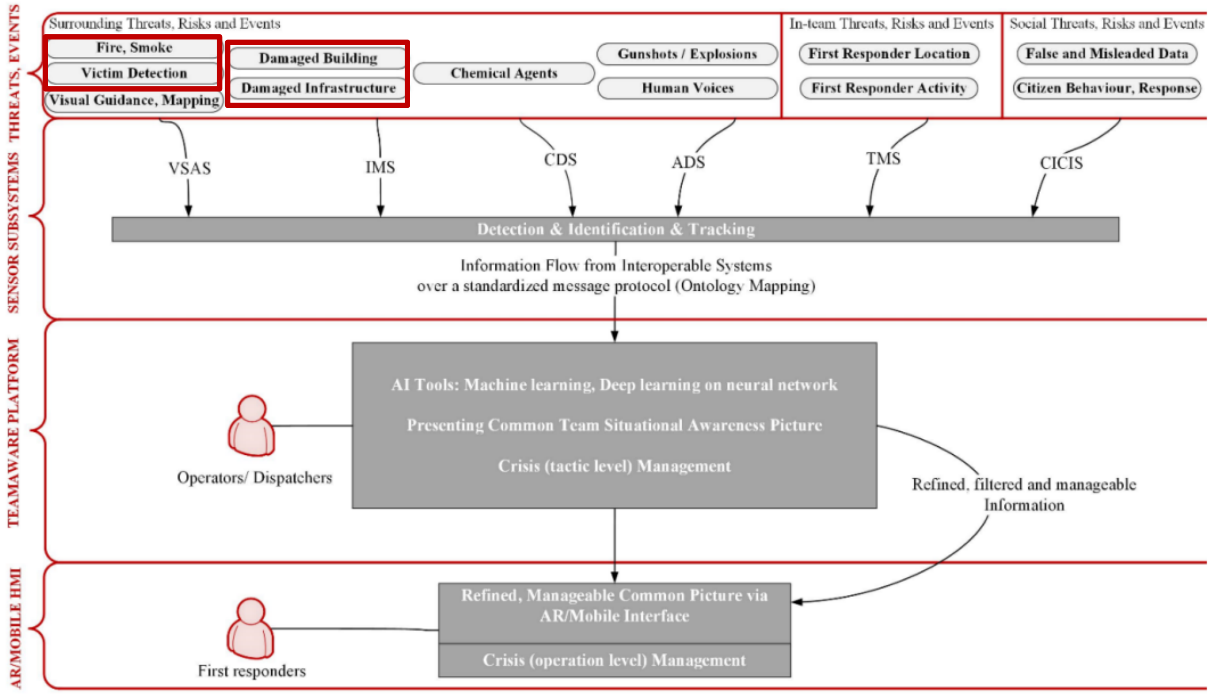

THREATS, EVENTS
Surrounding Threats, Risks and Events
Fire, Smoke
Victim Detection
Visual Guidance, Mapping
Damaged Building
Damaged Infrastructure
Chemical Agents
Gunshots / Explosions
Human Voices
In-team Threats, Risks and Events
First Responder Location
First Responder Activity
Social Threats, Risks and Events
False and Misleaded Data
Citizen Behaviour, Response
SENSOR SUBSYSTEMS
VSAS
IMS
CDS
ADS
TMS
CICIS
Detection & Identification & Tracking
Information Flow from Interoperable Systems
over a standardized message protocol (Ontology Mapping)
TEAMAWARE PLATFORM
Operators/ Dispatchers
AI Tools: Machine learning, Deep learning on neural network
Presenting Common Team Situational Awareness Picture
Crisis (tactic level) Management
Refined, filtered and manageable Information
AR/MOBILE HMI
Refined, Manageable Common Picture via AR/Mobile Interface
Crisis (operation level) Management
First responders

Figure 1. End users' operation in demonstration scenarios. The Red boxes present the main task the dataset of this report has to cover

The datasets, in this first phase of the project, are still a tool to start evaluating the various algorithms/approximations in the state of the art while waiting for the final datasets, which will be completed by the various end users and experts of the TeamAware project.

Future activities that will be set up concerns to merge appropriate datasets that comes out by the survey presented in this deliverable and acquisitions performed by the end-user in order to create a dataset adapted to First Responders field.

1.1 About This Deliverable

This first version of deliverable provides a survey on the most suitable datasets that can be deployed to solve WP3 tasks on computer vision applications. It will also put the bases of the procedure to follow for the data collection in the end-users' test bed.

A second version will be provided at it will take advantage of the element illustrated in this first, merging the entries of this survey, data coming from the end-users of the project and synthetic data.

1.2 Document Structure

The introduction and the executive summary described the needs of data to reach good quality performances for vision base inspection and situational awareness. In particular, these chapters stress the need of visual data in the field of First Responders applications. The introduction also presents the general procedure to follow to obtain new datasets in this domain leveraging, on well suited open-source datasets, pointing out the categories could be representative for visual recognition tasks in TeamAware project.

Chapter 2 presents different collection of public datasets devoted to victim detection. The aim is to use datasets commonly used for video surveillance, smart home assistance applications etc. As the problem of these datasets wants to solve is different from the victim detection, we are looking for only specific classes that we can adapt to the application.

Chapter 3 reviews datasets for image segmentation tasks for indoor and outdoor applications. The datasets presented have mainly an aerial point of view, as the TeamAware footages have a drone or a ceiling point of view coming from a CCTV system. The datasets illustrated in this deliverable have the task to map the element of the environment (indoor or outdoor) in an urban environment and detect fire or smoke.

Chapter 4 illustrates a dataset to detect building damages after an earthquake. This dataset merge data from earthquakes in Italy from 2009 to 2017. The aim is also to merge these data with a set of tailor-made data of artificial images, which is currently under development.

Chapter 5 describes the procedure put in place for the data collection in the end-users' test bed. The chapter presents the collection process, the description of the procedure and the element to highlight the composition of this future datasets.

Finally the conclusion part describes the work already performed and summarises the main important point to take into account in a second version of this deliverable.

1.3 Relation with Other Tasks and deliverables

This report is directly linked to the WP2 deliverables, looking to Figure 2, D3.1 is related to the scenario formulation elaborated in D2.4 and to D2.7 which presents the architecture and the solution selection of the overall TeamAware system. The data collection step of Figure 2 is an essential step to meet the objective of the scenario and their operational and functional requirements forecasted for the VSAS system.

The mapping between the VSAS system and operational requirements, developed in the context of Task 2.2 ("*End-users' needs, requirements, constraints and scenarios*"), is used as the main entry to support Task T3.1. The identification of the most relevant scenarios for severe disasters together with the end users of the technological partners involved in WP3 and WP4 providing the technological solutions for the visual assessment, as used as the support the reach the goals of this report.

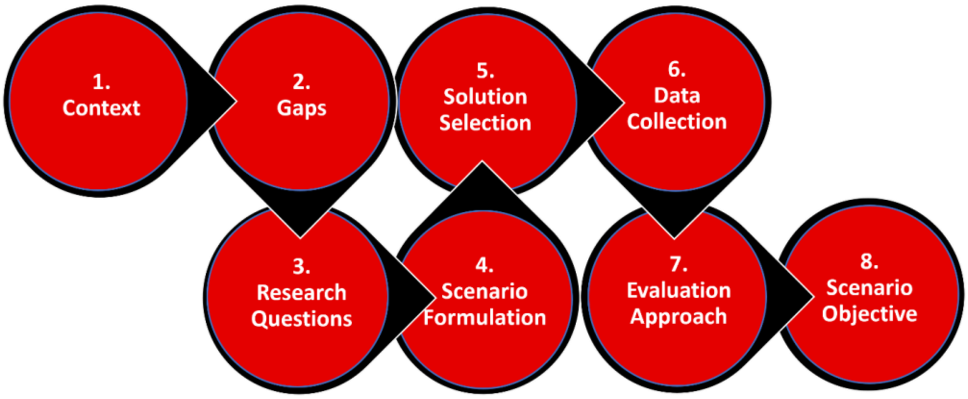


Figure 2 : TeamAware methodology and steps to reach the scenario objectives

2 Survey on Rescue Victim Datasets for Computer Vision applications

Deep learning is a popular method for human detection in the images and videos footprints, and it is applied extensively to the field of pedestrian detection, surveillance or smart home applications. The literature scarce for AI based applications in First Responders (FRs) and Law Enforcement Agency (LEAs) assistance fields. One reason behind this, is the unavailability of public datasets in this domain. As there is not a public dataset adapted to victim detection task, the aim is here to present various datasets, mainly used for human activity detection or pose estimation tasks, that could be filtered using only suitable categories and merge together in order to detect victims (i.e. people and animals) in an injured site.

The aim is to create a dataset for research and rescue applications, such as the tasks proposed in the TeamAware project, that has many defined categories in an attempt to cover all the possible situations and cases that the machine learning approaches (detectors, classifiers, Convolutional Neural Networks (CNN), etc.) may face while classifying the images sent from the injured site.

The categories we are looking for to detect victims include images of people in various poses: lying down, sitting, and falling etc. from different distances and different angles, where some images would show the whole human body, and others would show parts of the human body. In addition, there would be images captured in a controlled environment and “in the wild” conditions. Finally, the images containing other spaces and rooms with no human would be gathered.

2.1 Relevance to Project Scenarios

There are two distinct scenarios which are explained with details in D2.4 namely, the natural disaster and human-made disaster scenarios. VSAS will take part in both of the scenarios.

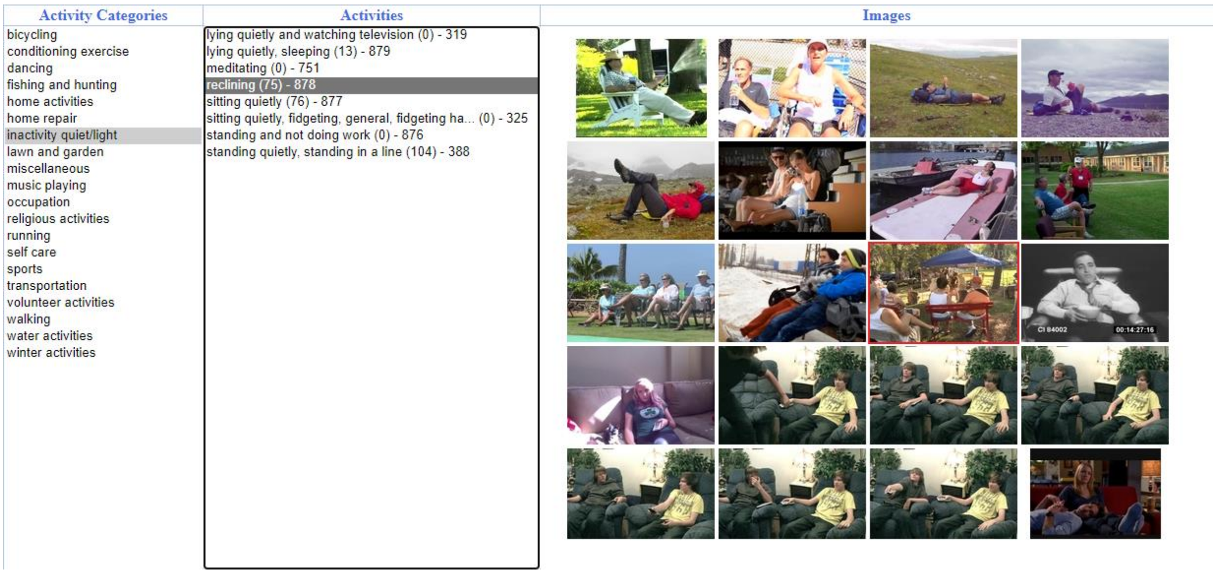
Considering the **natural disaster demonstration**, the VSAS system will be deployed in two phases:

1. A drone equipped with an RGB-D camera (Depth Sensor - RGB camera) will enter into the disaster zone for a first appreciation of the situation and to look for stranded passengers. The video footprints are sent and will be analysed into the command-and-control station. Thanks to the AI algorithms, previously trained with a victim detection dataset, a first evaluation of the victims will be performed.
2. After this first overview, an LEA wearing a helmet with an infrared (IR) camera and a standard grey camera will enter in the underground tunnel to rescue victims and to inspect in deeper the environment: other victims could be hidden by debris.

In the **human-made disaster scenario,** the helmet and the drone will be deployed and it will use to identify injured people in a demolished abandoned building

2.2 MPII Human Pose

Max-Planck Institute for Informatics (MPII) Human Pose Dataset, describe in [2], is a dataset for human pose estimation “in the wild”, in an indoor or outdoor context. It consists of around 25k images extracted from online videos. Each image contains one or more people, with over 40k people annotated in total. Among the 40k people samples, ~28k samples are for training and the remainder are for testing. Overall, the dataset covers 410 human activities (Figure 3) and each image is provided with an activity label. Images were extracted from a YouTube video and provided with preceding and following un-annotated frames.



Activity Categories

bicycling
conditioning exercise
dancing
fishing and hunting
home activities
home repair
inactivity quiet/light
lawn and garden
miscellaneous
music playing
occupation
religious activities
running
self care
sports
transportation
volunteer activities
walking
water activities
winter activities

Activities

lying quietly and watching television (0) - 319
lying quietly, sleeping (13) - 879
meditating (0) - 751
reclining (75) - 878
sitting quietly (76) - 877
sitting quietly, fidgeting, general, fidgeting ha... (0) - 325
standing and not doing work (0) - 876
standing quietly, standing in a line (104) - 388

Figure 3 : Example of the inactivity category in the MPII dataset that could be used for victim detections. Different categories are included in the dataset. In the framework of TeamAware project only few categories as laydown, sitting etc. will be considered for victim detection purpose

2.3 UAV-Human

Unmanned Aerial Vehicle (UAV)-Human, described in [3], is a large dataset for human behaviour understanding with UAVs It contains 67,428 multi-modal video sequences and 119 subjects for action recognition, 22,476 frames for pose estimation, 41,290 frames and 1,144 identities for person re-identification, and 22,263 frames for attribute recognition.

The dataset is interesting in the framework of TeamAware project because it presents data, which was collected by a flying UAV in multiple urban and rural districts in both daytime and night-time over three months, hence covering extensive diversities of subjects, backgrounds, illuminations, weathers, occlusions, camera motions, and UAV flying attitudes.

This dataset can be used for UAV-based human behaviour understanding, including action recognition (155 classes like sit down, stand up, move, walk…), pose estimation, re-identification, and attribute recognition. Furthermore, the drone is equipped with different sensors enabling the dataset to provide rich data modalities (Figure 4) including RGB, depth, IR, fisheye, night-vision, and skeleton sequences. This dataset fits very well to the sensors deployed on the TeamAware project, namely RGB-D, IR and standard grayscale camera in the of the CCTV system.

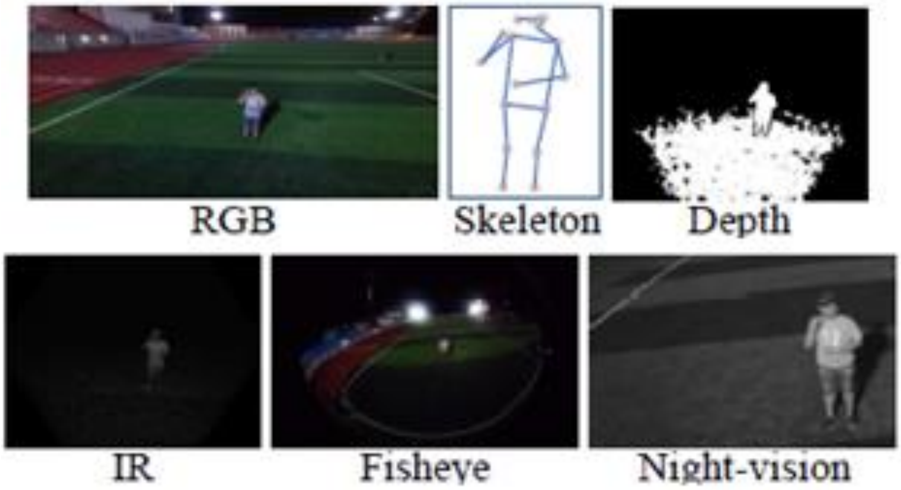


Figure 4 : Example of multisource data (multi modalities) using an UAV recorded in the UAV-Human Dataset

2.4 OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors

The dataset presented in “Opportunistic Passive Radar for Non-Cooperative Contextual Sensing” OPERAnet [4] is a collection of Indoor Video data obtained with two Kinects and RF measurements. OPERAnet is a comprehensive dataset developed with the aim to evaluate passive Human Activity Recognition (HAR) and localisation techniques with measurements obtained from synchronised Radio-Frequency (RF) devices and vision-based sensors.

The RF data includes Channel State Information (CSI) extracted from a Wi-Fi Network Interface Card (NIC), Passive Wi-Fi Radar (PWR) built upon a Software Defined Radio (SDR) platform, and Ultra-Wideband (UWB) signals acquired via commercial hardware. The vision/Infra-red based data are acquired from Kinect sensors.

Approximately 8 hours of annotated measurements are provided, which are collected across two rooms from 6 participants performing 6 daily activities. This dataset can be exploited to advance Wi-Fi and vision-based HAR, for example, using pattern recognition, skeletal representation, deep learning algorithms or other novel approaches to accurately recognise human activities.

Furthermore, it can potentially be used to passively track a human in an indoor environment. It is suggested to be used for development of new algorithms and methods in the context of smart homes, elderly care, and surveillance applications.

2.4.1 Kinect dataset description

Focusing on the Kinect dataset: the Kinect directory collects files are in “.mat” format and each row in the files corresponds to three-dimensional skeleton information captured from each of the two Kinects at a given point in time:

- exp_no: experiment number which is specified as "exp_002", "exp_003", etc. Note that the Kinect system does not need background scan. Hence, background data for "exp_001" and "exp_019" are omitted for the Kinect data.

- timestamp: UTC+01 00 timestamp in milliseconds when the Kinect skeleton data were recorded.
- activity: ground truth activity labels. The activity is specified as a string of characters with no spacing e.g., "walk", "sit", "stand", "liedown", "standfromlie", and “bodyrotate". These correspond to the activity numbers 1, 2, 3, 4, 5, 6 and 7 in the "Details" respectively.
- person_id: person ID specified as "One", "Two", "Three", etc.
- room_no: room ID specified as "1" (left room) or "2" (right room).

2.5 HMDB: a large human motion database

HMDB (Human Motion DataBase) presented in [5], is a collection of clips from various sources, mostly from movies, and a small proportion from public databases such as the Prelinger archive, YouTube and Google videos. The dataset contains 6849 clips divided into 51 action categories (Figure 5), each containing a minimum of 101 clips. The actions categories can be grouped in five types as general facial actions, facial actions with object manipulation, general body movements including sit down, sit up, etc. body movements with object interaction, body movements for human interaction.

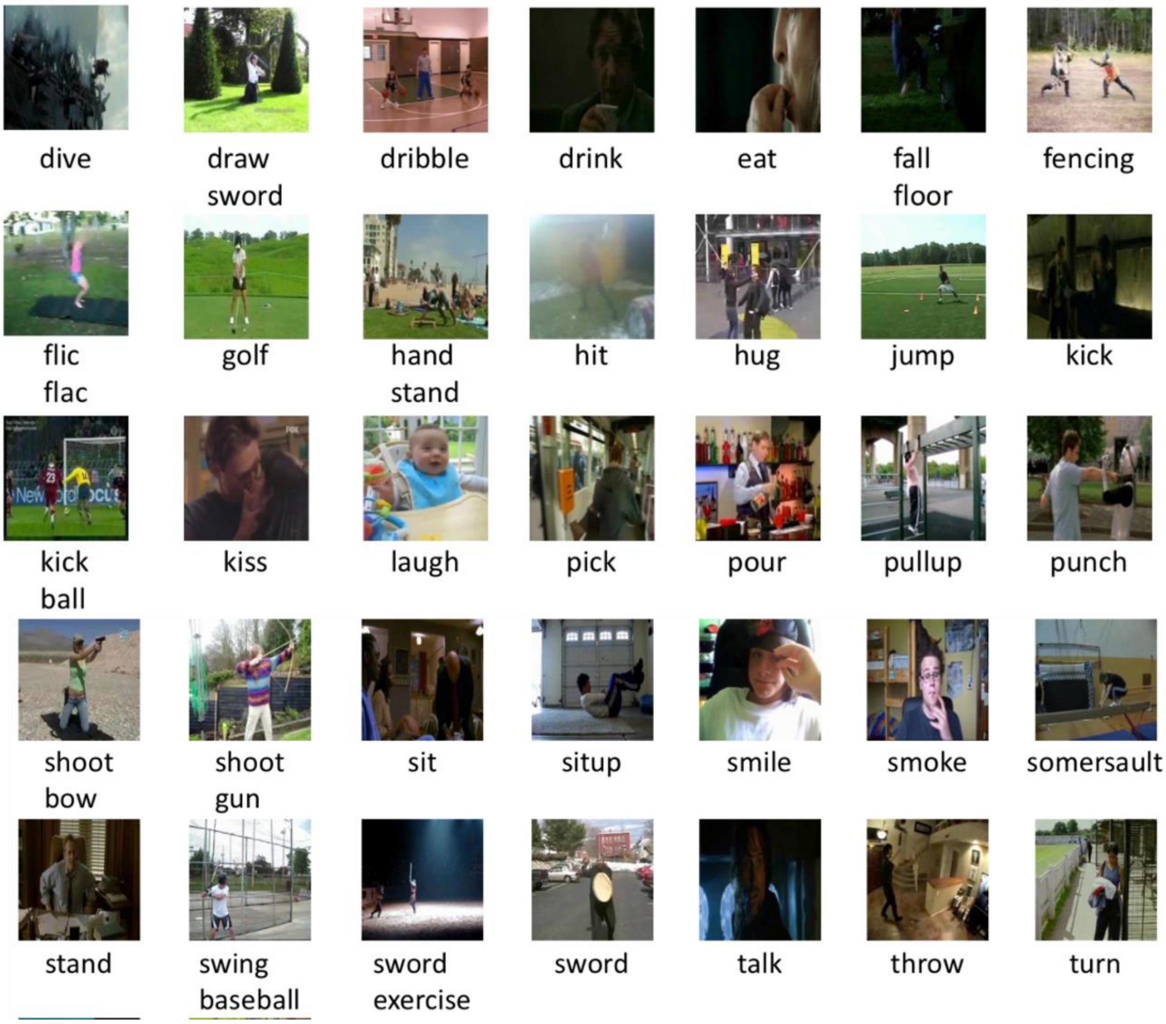


Figure 5 : Sample of data from HMBD: fall floor and sit categories could be used in the TeamAware project

2.6 Hollywood 3D: Recognising Actions in 3D Natural Scenes

The paper presented at IEEE conference in Computer Vision [6], focus on action recognition in unconstrained situations. This is a 3D dataset for action recognition in the wild. The detection and recognition of actions in natural settings is useful in several applications (Figure 6), including automatic video indexing and search, surveillance and assisted living.

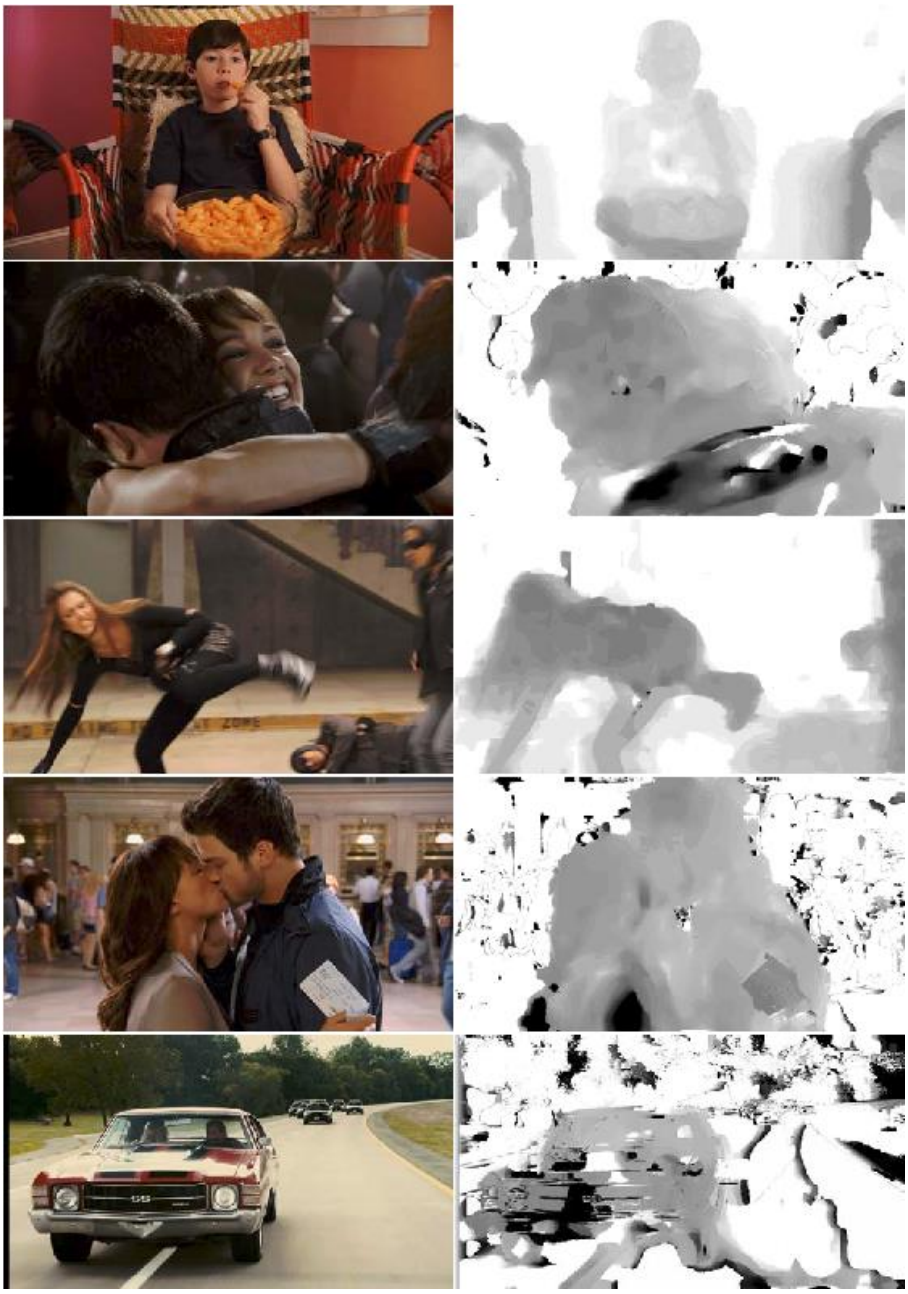
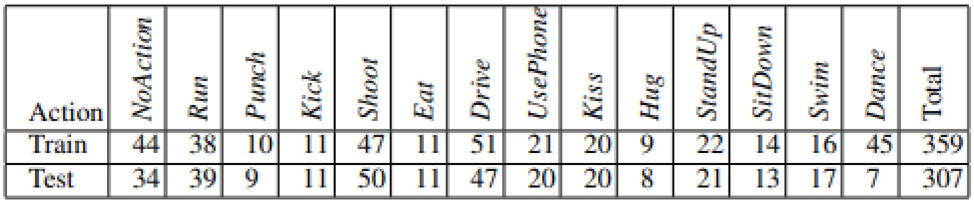


Figure 6 : Sample of data from Hollywood3D with the correspondent depth maps

This dataset presents a new level of complexity to the recognition community, arising from the natural within-class variation of unconstrained data, including unknown camera motion, viewpoint, lighting, background and actors, and variations in action scale, duration, style and number of participants. While this natural variability is one of the strengths of the data, the lack of structure or constraints make classification an extremely challenging task.

The Table 1 below shows the number of training and test sequences available for each action in the dataset, ensuring separate films are used for training and test data.

Table 1: Number of test and train sequences for each category in the Hollywood3D Dataset.



Action	*NoAction*	*Run*	*Punch*	*Kick*	*Shoot*	*Eat*	*Drive*	*UsePhone*	*Kiss*	*Hug*	*StandUp*	*SitDown*	*Swim*	*Dance*	Total
Train	44	38	10	11	47	11	51	21	20	9	22	14	16	45	359
Test	34	39	9	11	50	11	47	20	20	8	21	13	17	7	307

2.7 MSRDailyActivity3D

This dataset was introduced by Jiang Wang et al. in mining action let ensemble for action recognition with depth cameras [7]. DailyActivity3D dataset is a daily activity dataset captured by a Kinect device. There are 16 activity types (Figure 7): drink, eat, read book, call cell phone, write on a paper, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lay down on sofa, walk, play guitar, stand up, sit down. If possible, each subject performs an activity in two different poses: “sitting on sofa” and “standing”.

The total number of the activity samples is 320. This dataset is designed to cover human’s daily activities in the living room. When the performer stands close to the sofa or sits on the sofa, the 3D joint positions extracted by the skeleton tracker are very noisy. Moreover, most of the activities involve the humans-object interactions.

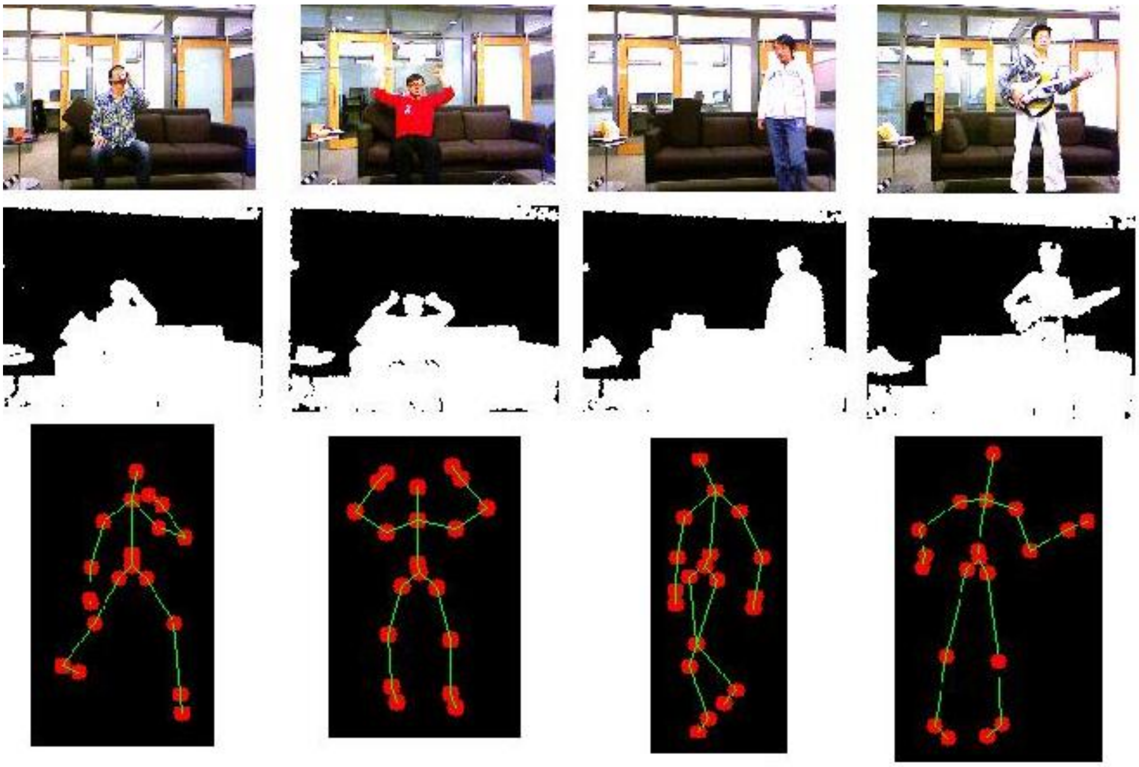


Figure 7 : Data sample with RGB, with related depth maps and skeleton sequences

2.8 Summary of Dataset for Victim Detection

This chapter presents several datasets that will be useful for the TeamAware project. Due to the diversity of sensors (RGB, depth, IR, skeleton information) and the diversity of points of view (frontal, human centric, UAV…), these datasets constitute a first varied and relevant dataset to develop and train models based on the human detection and people pose estimation**.**

After this significant part centred on finding data from databases and publicly available repositories that can help to solve the victim detection problem, it will be necessary to clean the data frames, work with multi-dimensional arrays, and manipulate data frames to aggregate data.

To deploy this data in the TeamAware framework, it will be necessary to building an efficient data architecture, streamlining data processing, and maintaining large-scale data systems.

Beside to that the road map for the deployment of the data imply the definition of the Deep Learning algorithm suitable for the analysis of this data.

3 Survey on Images Segmentation Datasets for VSAS

3.1 Switzerland Aerial Drone Footage Datasets

This dataset can be found at [8]. It contains 4K-quality videos captured with a DJI Mavic Pro drone [9] under different conditions, which are representative of real drone footage in terms of contents and camera angles. The videos have minimal post-processing (colour correction) and do not include any ground truth information, so they would have to be annotated, possibly including multiple object classes in each frame so that they can be used to train different detectors. Although the rich contents in these videos will provide valuable information for general object detection, training data will still be required for scenarios with more specific particularities.

3.1.1 Relevance to Project Scenarios

The dataset contains 13 bird-eye drone videos from natural scenery (snow mountains, forests, sea, roads, urban, etc.) with and without people in it, which can be useful to train reconnaissance and environmental mapping algorithms regarding the natural disaster scenario. Videos contain elements such as trees, mountains, rocks, birds, cars, etc., which can be used to train detectors for different elements (both natural and man-made). Figure 8 shows thumbnails from all the videos included in this dataset, where the relevance of its contents can be seen.

In case of an attack regarding human-made disaster, the first task to be performed is to evaluate the situation. Usually, photographs of the scene are required by analysts to perform such evaluations. In TeamAware, the use of drones, CCTV systems, and head mounted cameras will provide this information, so automatic computer vision algorithms must be able to analyse and map the environment before decisions can be taken.

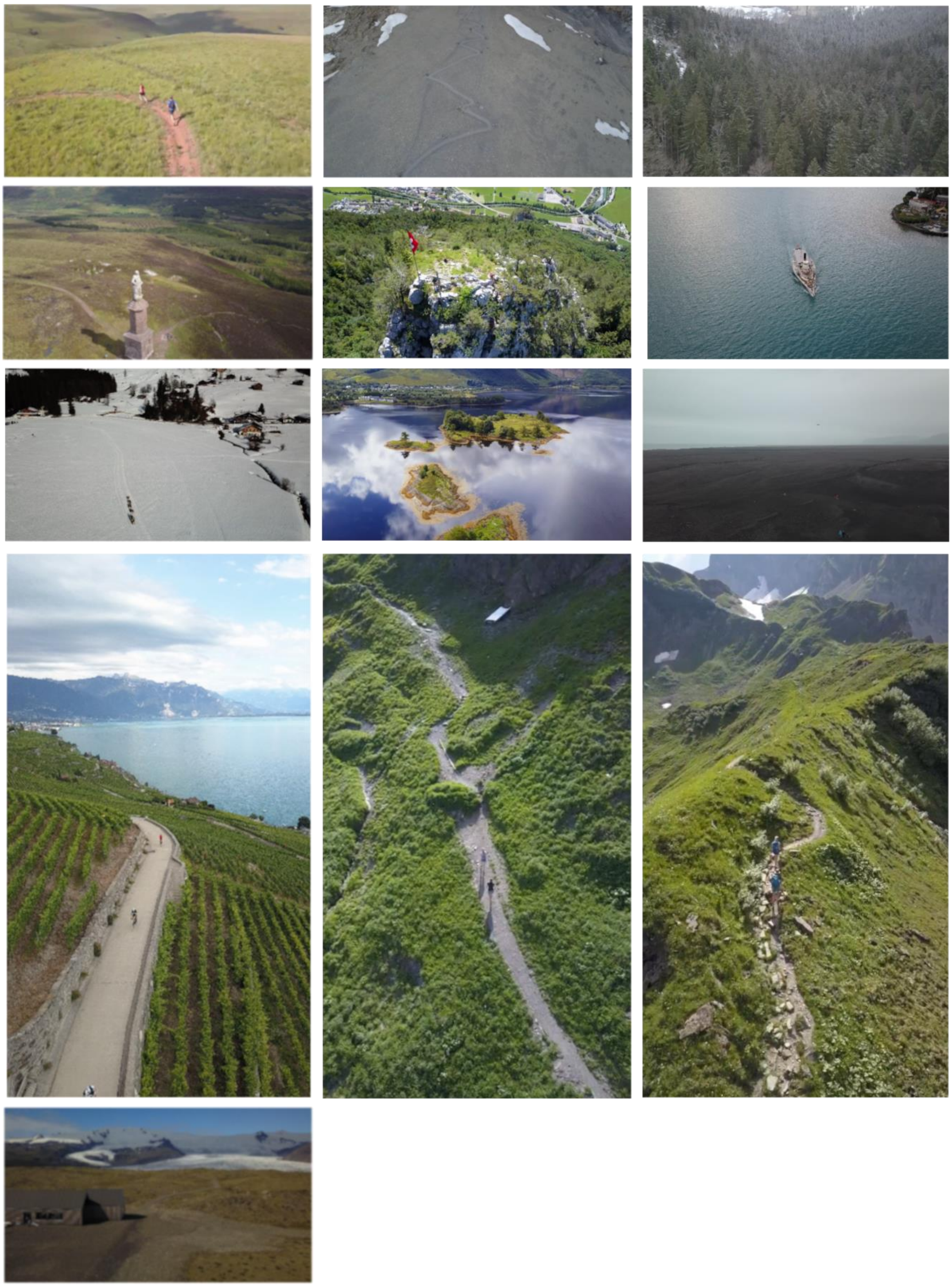


Figure 8: Switzerland Aerial Drone footage dataset thumbnails

3.2 USTC Forest Smoke and Fire Dataset

This dataset has been created by the University of Science and Technology of China (USTC) and it's specific for smoke and fire detection tasks with colour cameras. The full dataset is available at [9] and it requires authentication, but a reduced version has been made publicly available on Kaggle at [10].

The reduced version contains 15.800 images, but no ground truth data is provided, which means that all these images would have to be annotated.

The images are divided into four different sets:

- training sets for smoke detection (Figure 9) and fire detection (Figure 10) tasks
- 2 test sets (one large and one small) containing mixed fire and smoke images.

3.2.1 Relevance to Project Scenarios

The detection of fire and smoke is a task described in DOA and also requested by the end users, and USTC dataset can be a very good starting point to train the detection algorithms on. Figure 11 and Figure 12 illustrate some of the contents in the dataset, merging smoke and fire images for small and big test on the segmentation algorithms.

Figure 9: USTC Forest smoke and fire dataset. Smoke training set preview

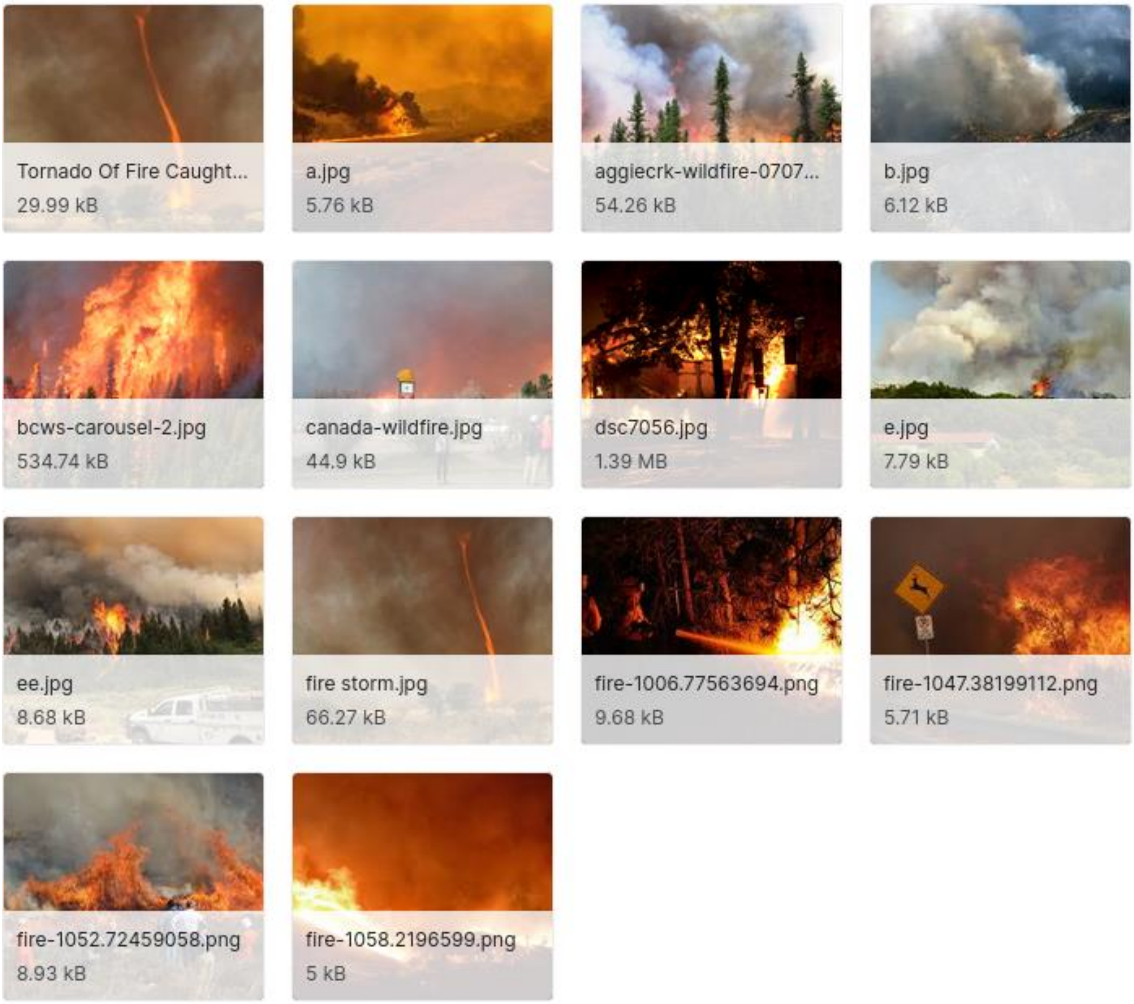
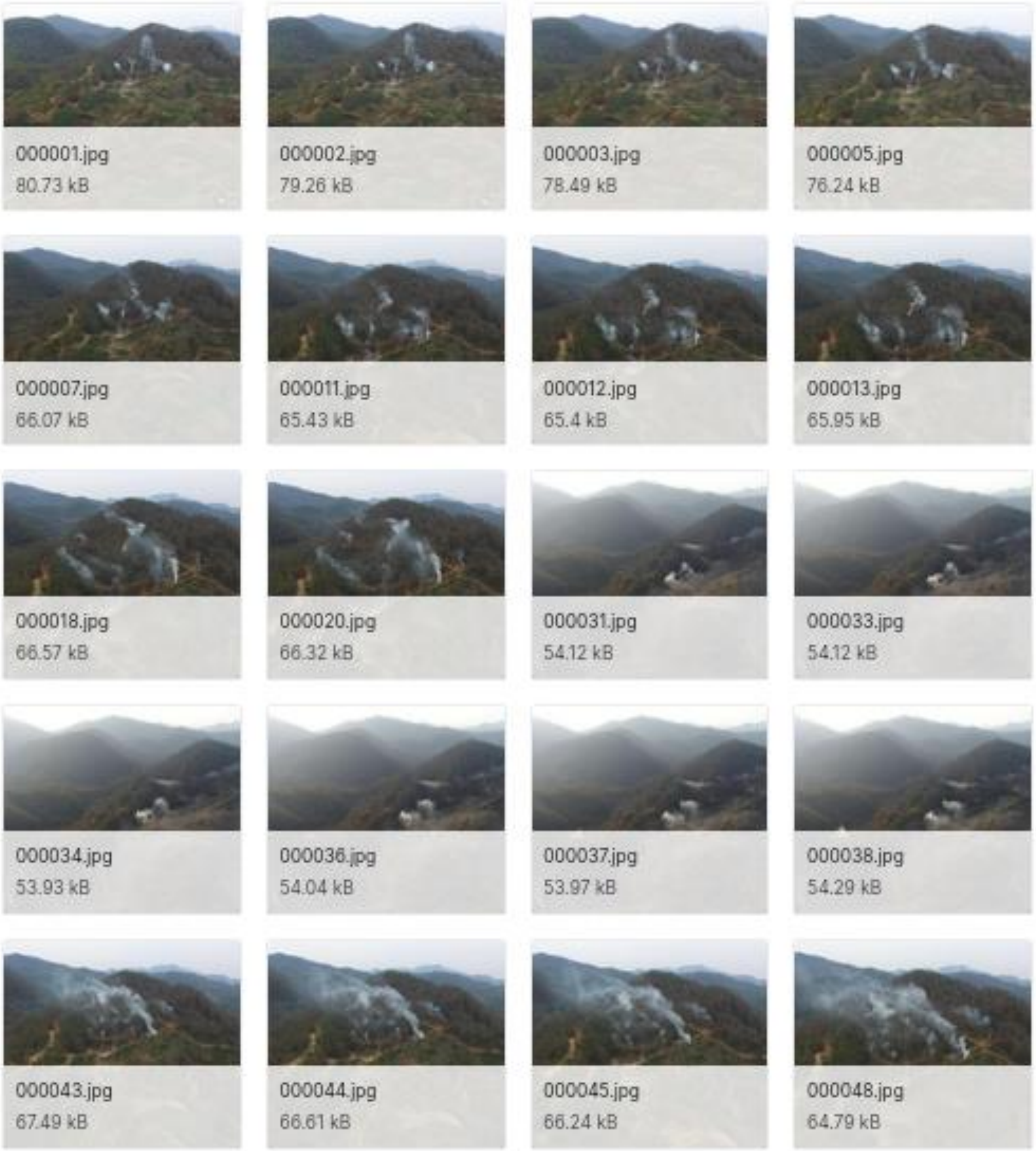


Figure 10: USTC Forest smoke and fire dataset. Fire training set preview



000001.jpg
80.73 kB

000002.jpg
79.26 kB

000003.jpg
78.49 kB

000005.jpg
76.24 kB

000007.jpg
66.07 kB

000011.jpg
65.43 kB

000012.jpg
65.4 kB

000013.jpg
65.95 kB

000018.jpg
66.57 kB

000020.jpg
66.32 kB

000031.jpg
54.12 kB

000033.jpg
54.12 kB

000034.jpg
53.93 kB

000036.jpg
54.04 kB

000037.jpg
53.97 kB

000038.jpg
54.29 kB

000043.jpg
67.49 kB

000044.jpg
66.61 kB

000045.jpg
66.24 kB

000048.jpg
64.79 kB

Figure 11: USTC Forest smoke and Fire dataset. “Big” test set preview

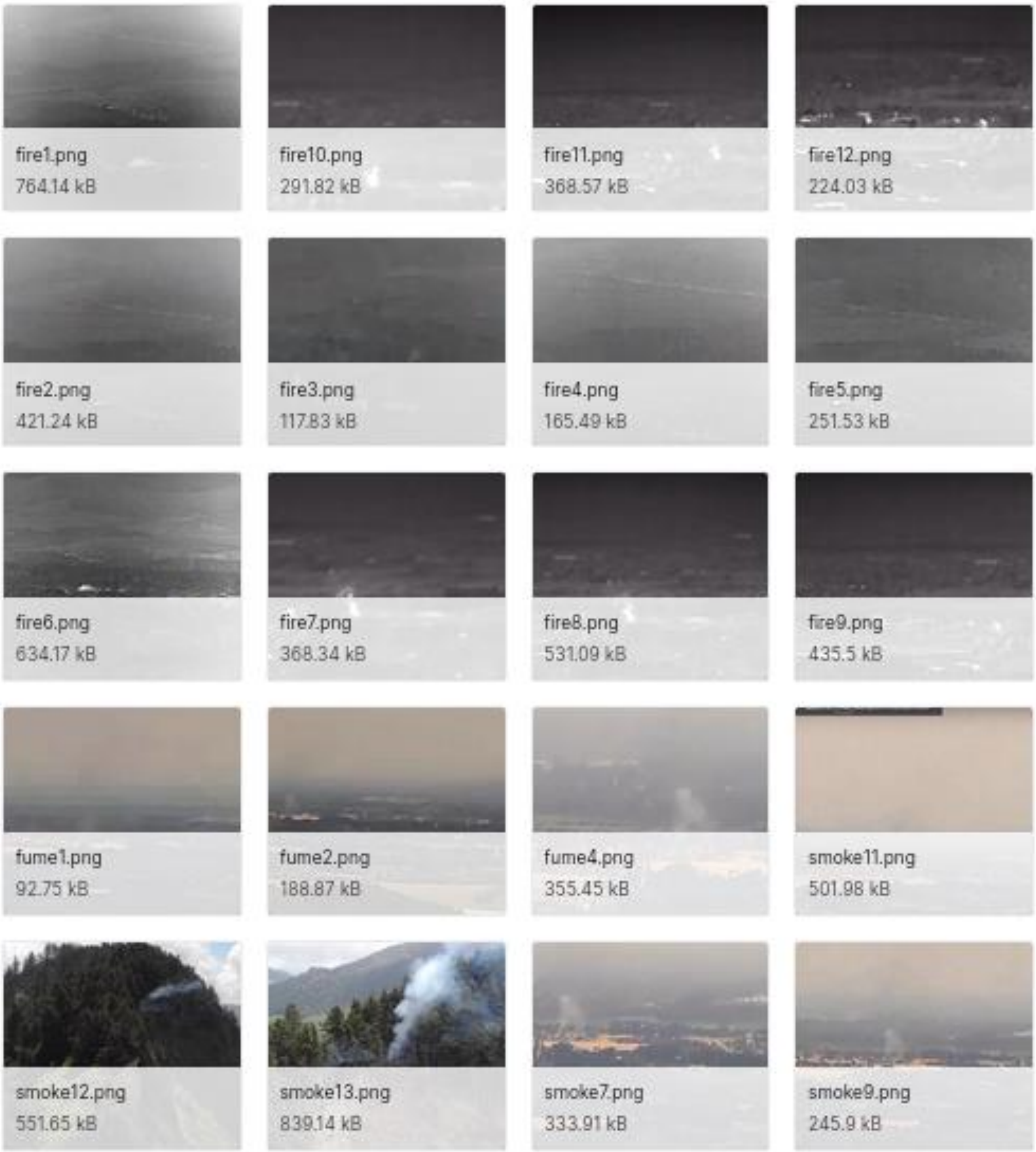


Figure 12: USTC Forest smoke and Fire dataset. “Small” test set preview

3.3 The Stanford Drone Dataset

This dataset has been created by the computational vision and geometry lab at Stanford University. The dataset consists of annotated videos of pedestrians, bikers, skateboarders, cars, buses, and golf carts navigating eight unique scenes on the Stanford University campus. It was created to analyse the behaviour of different types of agents when crossing paths in a real-world outdoor environment, to improve tasks like target tracking or trajectory forecasting. This dataset contains ground truth annotations which simplifies its use in TeamAware.

The original 69GB dataset can be found at [11], although a full compressed and optimised version of only 2GB is available on Kaggle [12]

3.3.1 Relevance to Project Scenarios

Although the original purpose of this dataset (a sample is shown in Figure 13) may not be of much direct use to the TeamAware project, the high number of buildings, streets, and other natural and structural elements found in many public spaces can be of great use for mapping and detection tasks in both of the project scenarios, especially for cases when disasters occur in outdoor public spaces like the ones depicted in this dataset. The downside to this dataset is that, to use it for training on different classes than the ones annotated in it will require additional annotation work.

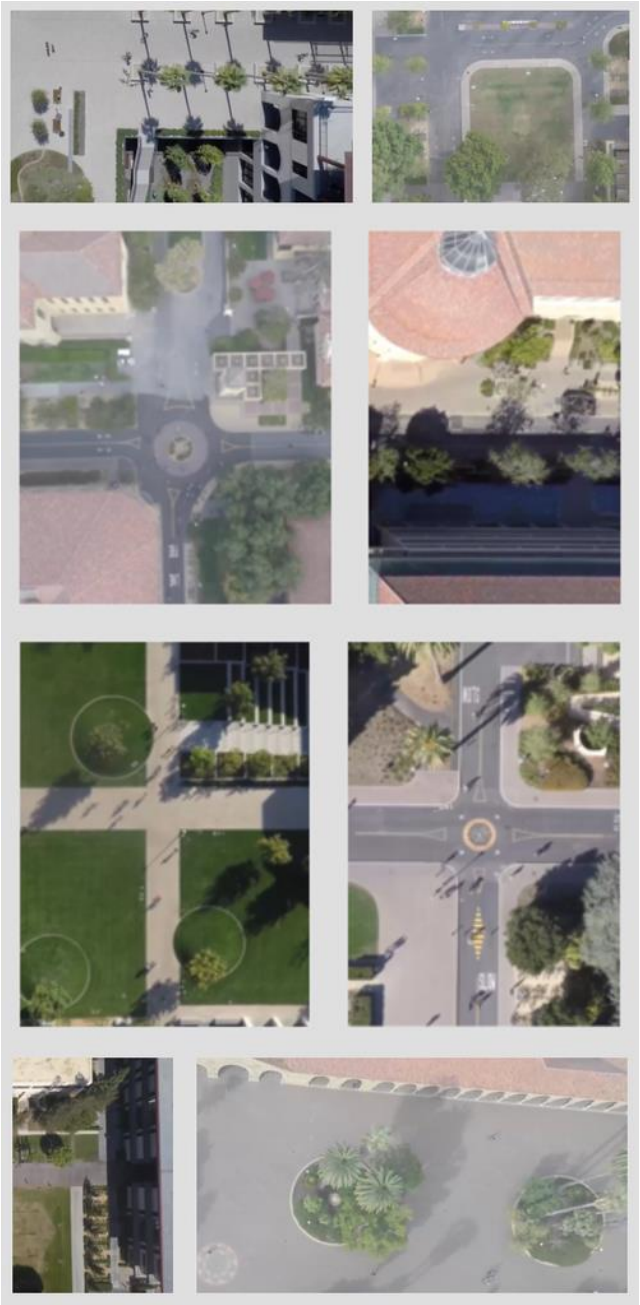


Figure 13: Stanford drone dataset preview

3.4 Aerial Dataset of Floating Objects (AFO) Dataset

This is a dataset of aerial images for maritime search and rescue applications [13]. With over 40,000 hand-annotated elements in aerial-drone videos, this is the first free dataset of this type for training ML/DL models. It contains images from fifty video clips depicting objects floating on the water surface, captured by different drone-mounted cameras of different resolutions. The dataset has 3.647 images with 39.991 annotated objects, split into three parts: a training set (67.4%), a test set (19.12%), and a validation set (13.48%), where the test set is taken from specific unseen videos to avoid overfitting.

Attention must be paid to the license of this work, as the dataset is published under the Creative Commons Attribution-Non-commercial-Share Alike 3.0 License and so it cannot be used for commercial purposes or without attribution.

3.4.1 Relevance to Project Scenarios

This dataset is a useful resource to train algorithm in the detection of people in different positions like lying face up and face down, sitting, swimming, submerged from the neck down, etc. This could have applications not only in water scenarios but also in other different situations on the ground, both indoors and outdoors. Many of the images show subjects in groups, which can also be used to detect clusters of people in rescue operations, or to detect anomalous behaviour in terrorist scenarios.

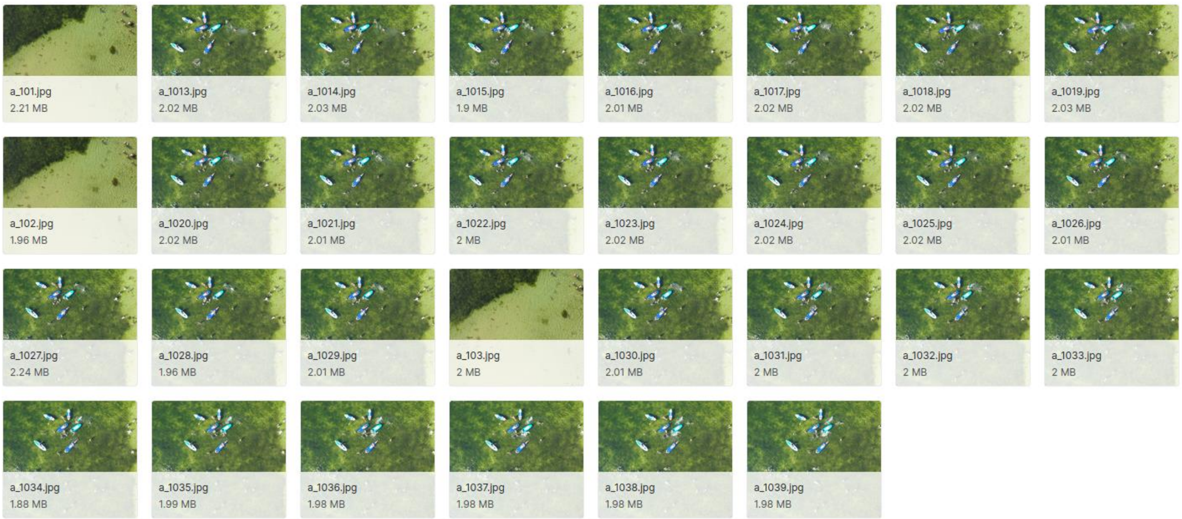


Figure 14: Aerial dataset of floating objects. Set 1 preview

Figure 15: Aerial dataset of floating objects. Set 2 preview

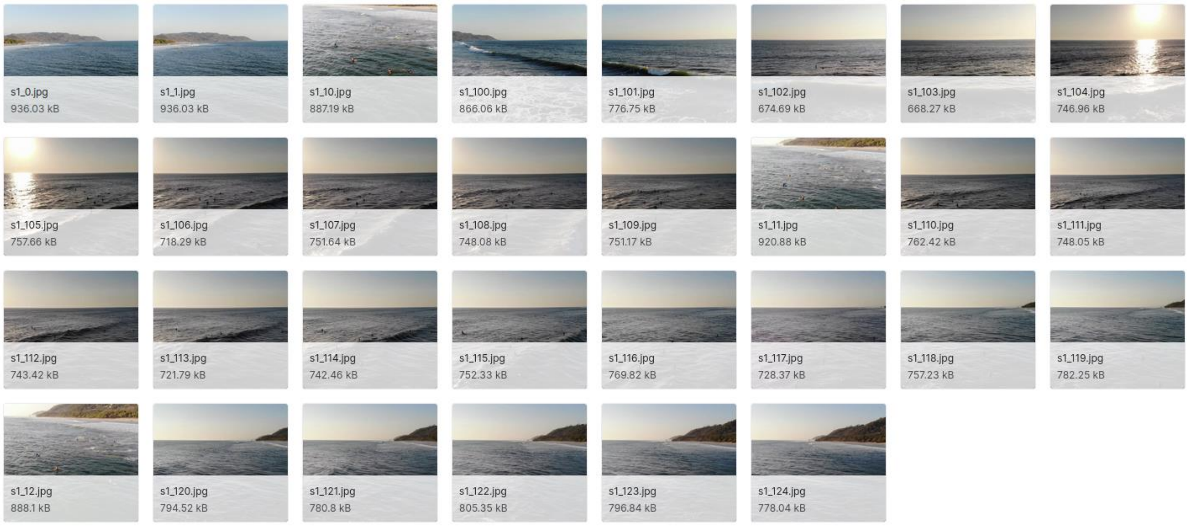


Figure 16: Aerial dataset of floating objects. Set 3 preview

3.5 Low Altitude Disaster Imagery (LADI) Dataset

The LADI dataset provides images that focuses on the Atlantic Hurricane and spring flooding seasons since 2015. Two key distinctions are the low altitude, oblique perspective of the imagery and disaster-related features, which (according to the authors) are rarely featured in computer vision benchmarks and datasets. The dataset uses a hierarchical labelling scheme of a five coarse categorical and then more specific annotations for each category. The five coarse categories are:

- Damage

- Environment
- Infrastructure
- Vehicles
- Water

For each of the coarse categories, there are 4-9 more specific annotations (Table 2).

Table 2: Hierarchical labels use in the annotation of LADI dataset.

Damage	Environment	Infrastructure	Vehicles	Water
damage (misc.)	dirt	bridge	aircraft	Flooding
flooding / water damage	grass	building	boat	lake / pond
Landslide	lava	dam / levee	car	Ocean
road washout	rocks	pipes	truck	Puddle
rubble / debris	sand	utility or power lines / electric towers		river / stream
smoke / fire	shrubs	railway		
	snow / ice	road		
	trees	water tower		
		wireless / radio communication towers		

Information and documentation on the dataset can be found on GitHub [14] as well as the dataset's page on AWS's open data registry [15]. The full dataset (327,2GB) can be downloaded from the project's AWS S3 bucket at [16].

Figure 17: Sample images from the LADI dataset

3.5.1 Relevance to project scenarios

The applications of this vast dataset to the TeamAware projects are numerous. From the detection of several landmarks (lakes and other bodies of water, green areas, forests, etc.), houses, mountains, etc., to locating landing sites for responders, roads, tracks, access points and many other areas that can be of interest during search and rescue operations.

3.6 Indoor Semantic Dataset (2D-3D segmentation)

The dataset [17] includes semantic and geometric data in 2D, 2.5D, and 3D domains, as well as their instance-level annotations. The dataset consists of about 70,000 RGB images, along with the corresponding depths, surface normal, semantic annotations, global XYZ images, as well as camera information. Apart from these images and information, there are also point clouds and raw 3D meshes registered and semantically annotated. In itself, this dataset allows the development of joint and intermodal learning models and potentially unsupervised approaches that use the regularities present in large-scale interior spaces.

3.6.1 Relevance to project scenarios

This dataset collects on 6 large-scale indoor areas originating from 3 different buildings of primarily office and educational use. These data are clearly useful in any of the two scenarios proposed by the project (natural or man-made disaster), since in both cases, it is necessary to recognise the interior spaces of buildings to plan necessary rescues.

The contained data is displayed in two modes: 3D and 2D. In 3D mode, the data contains coloured point clouds and textured meshes for each scanned area. 3D semantic annotations for objects and scenes are provided for both modalities, with corresponding point-level and face-level labels. Annotations were initially done on the point cloud and projected onto the surfaces of the models, as well as indicating the bounding boxes of the objects in the scene. In the 2D modality, the data included is the RGB and depth raw images in full high definition with a resolution of 1080x1080, as well as their annotations of the objects present in them. Examples of these annotations can be seen in Figure 18 (2D data) and Figure 19 (3D data).

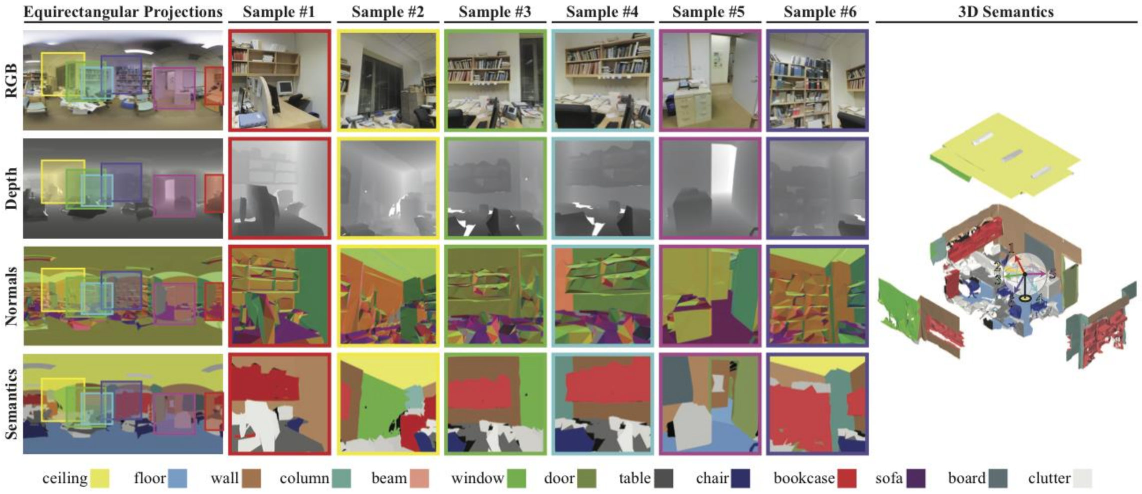


Figure 18: Sample images from the 2D modalities catalogue of the dataset.

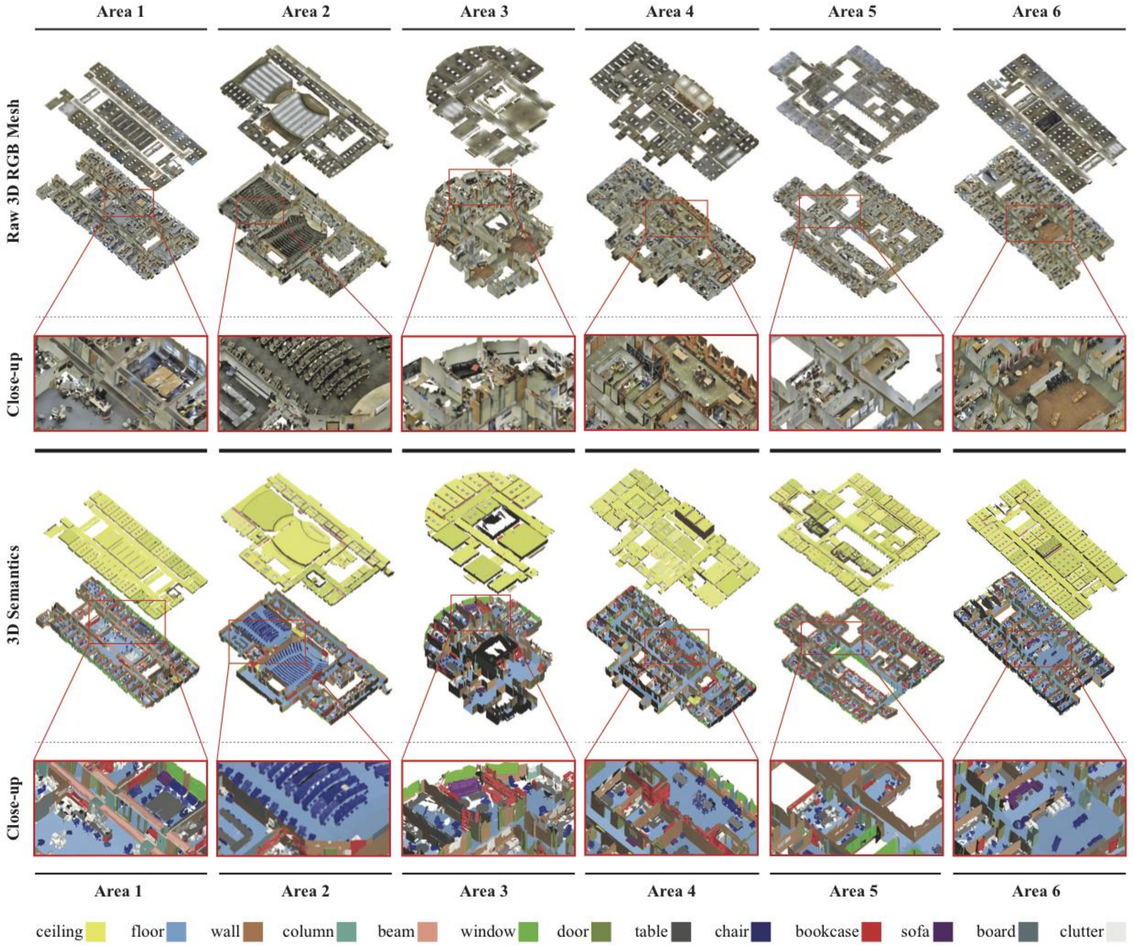


Figure 19: Dataset in 3D modalities includes: the textured 3D mesh models, semantic annotation and their point clouds

3.7 Amazon Web Service Open Data Registry

The Amazon Web Service (AWS) Open data registry hosts around 320 searchable datasets for specific tasks, 45 out of which are specific to disaster recovery alone. Most of these datasets provide landmark satellite imagery of different types, mostly used for agriculture purposes, which could be useful for certain detection and mapping tasks in TeamAware.

3.8 Summary of Image Segmentation Datasets for VSAS

This chapter presents a series of datasets that can be useful to train some of the TeamAware project's detection algorithms, for the sample scenarios proposed in the project as well as others that may come up later on. The datasets provide valuable resources to develop and train deep learning algorithms for detection and recognition tasks from drones and wearable cameras, which are the main source of video footage that will be used in the project. Although some of these datasets require additional annotation work (which could potentially extend the duration of some of the tasks), the benefits of having so much data (especially in cases where training data is scarce) could justify this additional work. Nonetheless, despite having these datasets, data for specific objects, tasks, and scenarios is still very much needed.

4 Survey on Building Defect Datasets

4.1 Relevance to Project Scenarios

Photos of real damages after catastrophic events and synthetic images of damaged structures constitute the datasets on building/infrastructure damage owned and implemented by EUCENTRE. Such datasets will be deployed in TeamAware for training and internal validation processes for the implementation of the AI algorithm for damage detection, which is at the base of the Infrastructure Monitoring System (IMS), as described in deliverables D2.7 and D4.1. These datasets, in particular the one composed by photos of reals cases, are constituted by annotated images, in which several structural elements and structural damages are identified. For both datasets, the event of reference is mostly a natural disaster, in particular earthquakes but also ground deformations (such as landslides, settlements or subsidence) that may jeopardise the structural safety. The use of the datasets may also be extended to those manmade disasters that could cause damages to structural components or to those conditions in which a suitable maintenance level for infrastructure operability is not guaranteed.

4.2 EUCENTRE’s Dataset of Post-event Damages and Structural Inspections

The EUCENTRE’s dataset collects photos from surveys performed on different occasions:

- Post-event structural safety assessments after the last three major seismic sequences in Italy: the L'Aquila earthquake (2009), the Emilia earthquake (2012) and the Central Italy earthquake (2016-2017). A sample is shown in Figure 20;
- Structural inspections on infrastructures to monitor their maintenance level and structural stability.

The dataset, constantly increased and updated, is at the time of writing constituted by approximately 40000 images of which the greatest part regards post event seismic assessments. Each image collected undergoes a process of annotation, based first on the identification of the structural element and subsequently of the damage (whose classification depends on the structural element typology). There are two particular cases in which the damage is not associated with a single element: the first may be a phenomenon globally affecting the structure (e.g., total collapse), the second could be a single type of damage associated with non-structural elements. Besides information of structural characteristics and damage level, images are annotated taking into consideration the geographical location, the use of the structure, the period in which the survey was performed.

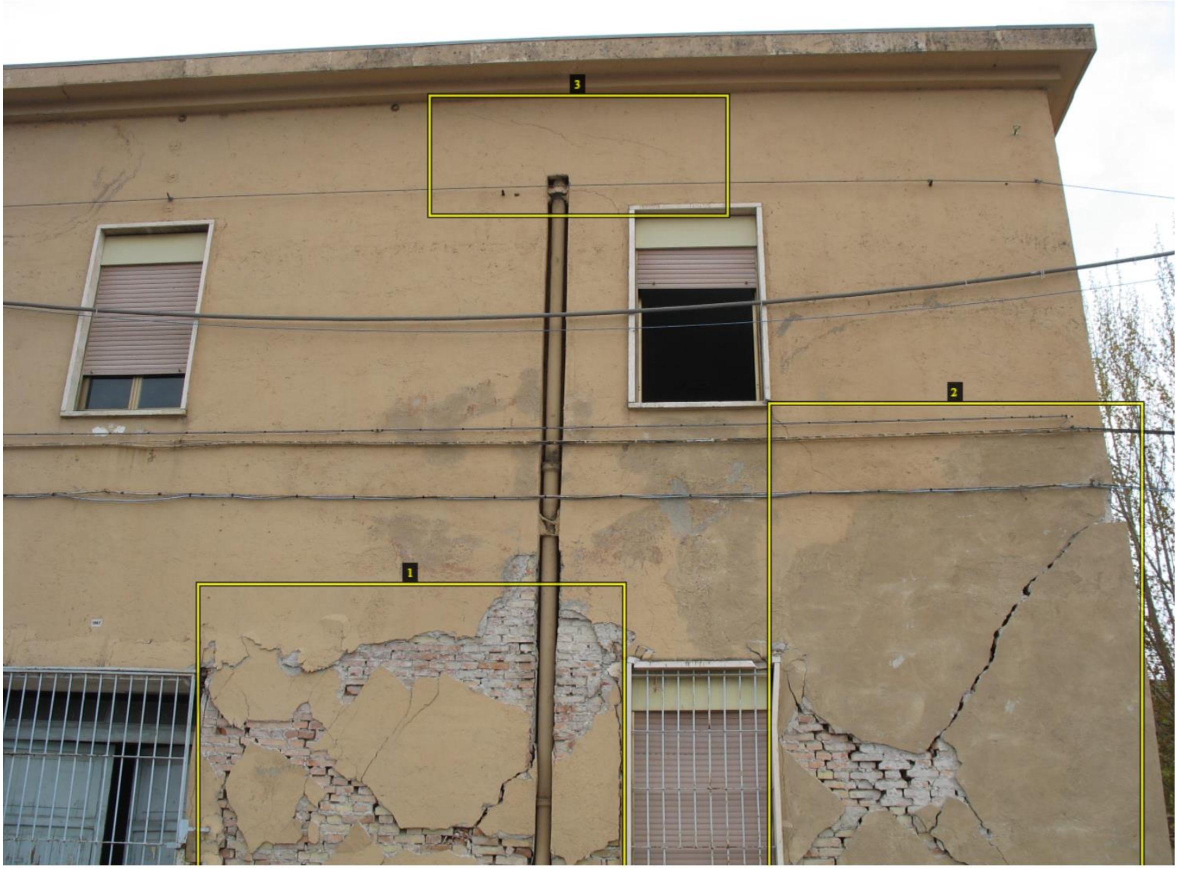


Figure 20: Example of an annotated image belonging to EUCENTRE'S dataset (L'Aquila 2009 event, shear damage on unreinforced masonry piers and spandrels)

Most part of the photos were taken with typical RGB cameras, with camera sensors of different image resolutions (e.g., Canon EOS 400D).

Some of the inspections of infrastructures (bridges) for maintenance monitoring were performed with optical RGB drones:

- DJI Spark: 1/2.3” CMOS, Effective pixels: 12 MP, Lens35 mm Format Equivalent, aperture: f/2.8, image size 3968×2976;
- DJI Mavic 2 PRO: 1'' CMOS, Effective Pixels: 20 MP, Lens: 35 mm Format Equivalent, aperture: f/2.8-11, image size 5472×3648;
- Anafi Parrot: 1/2.4’’ CMOS, Effective Pixels: 21 MP, Lens35 mm Format Equivalent, image size 5344x4016.

The environment in which the photos are shot is typically outdoor, with good visibility, usually in daylight with sun or light clouds, rarely rainy or severely foggy. A limited sample of images is shot indoor, again in good visibility conditions (the number of those images is limited to few hundred samples at most).

The point of view in general is at ground level for the post event surveys after earthquakes, although few hundred samples of photos of bridges are characterised by an aerial point of view, since they were taken with drones.

The dataset belongs to the EUCENTRE Foundation, with all rights reserved. Access to the dataset may be allowed upon request and for research purposes only with appropriate citation.

4.3 Report of the Synthetic Dataset for Building Defect Recognition by EUCENTRE

Deep Neural Networks are the state of the art of most of object detection algorithms. Such method, however, needs a significant number of examples of each object class to be detected (in the order of several thousands) in order to be sufficiently robust for industrial applications. This constraint is particularly difficult to overcome when the target object classes are attributable to anomalies of some kind, which, in the case of structural/infrastructural damages, may be rare hence determining datasets of limited volume. The motivation for which the EUCENTRE Foundation is creating a tailor-made dataset of artificial images is because the synthetic generation ones provides images of sufficiently comparable characteristics to real ones, in support to the operations of training and validation of object detection algorithms.

The synthetic dataset, at the date of editing of the present deliverable, is currently under development.

The images that will be included in the dataset will represent the following:

- A subset of the most common building/infrastructure typologies (e.g., residential buildings, reinforced concrete girder bridges, unreinforced masonry churches, etc.).
- A subset of the most common structural damages aimed to represent the effect of both natural catastrophes (in particular earthquakes, landslides in which the ground deformation affects the structural stability) or lack of maintenance and weathering (i.e. cracks, spalling, leaching and corrosion are included).

The dataset is created starting from 3D point cloud or 3D mesh models obtained from aerial surveys with drones of real structures. The images of damages to train the Object Detection algorithm at the base of TeamAware's Infrastructure Monitoring System (IMS) will be produced and extracted thanks to the use of the software Blender, considering as a pre-condition that the simulated survey is conducted only outdoors in daylight and with good visibility, by varying the following:

- point of view from which the image is shot (either reproducing a survey performed with a typical photo camera at ground level or an aerial survey with a drone at different heights and distances from the structure);
- light and exposure conditions, by varying luminosity (for sun, clouds, time of the day simulation), presence and intensity of shades;
- texture and colour of structural elements;
- typology, level of severity and location of the selected typologies of damages.

4.4 Summary of Building Defect Datasets

This chapter presents the two datasets that will be useful for the training and validation of the IMS in TeamAware's platform. The datasets, one constituted by real-word images, the other by artificially generated ones, are representative of several classes of damages, deemed relevant for structural safety assessment.

5 Requirements for the on the In-site Recording Session

This section presents the procedure and the related set of questions covering the information on the collection process and the scenario the data will cover. This work will also provide the starting point of the V2 of deliverable D3.1 describing the datasheet useful to dataset consumers in the framework of TeamAware WP3 and relevant links to WP4.

5.1 Structure of Data Collection Process

This preparation phase, called "Structure of data collection process", has 3 main objectives listed here:

1. Motivation *(i.e. "Hidden Victim Detection" - Help first responders wearing the helmet with the IR camera to locate potential victims that the Drone – equipped with an RGB-D camera - in a first inspection couldn't detect because victims are hidden under the debris)* and the associated questions to describe to task the data collection will *cover.*
2. The dataset composition (*i.e., how many data, how many data with partially occluded human body, how many data presenting bodies entirely occluded, etc.)* and the associated questions.
3. The description of the Collection process starting from physical sensors, participants etc.…

5.1.1 Motivation

Here are listed the main question related to the motivation part. The datasets must be clarified the following items:

- For what purpose was the dataset created?
- Which is the associated scenario?
- Was there a specific task in mind? *(if necessary) Was there a specific gap that needed to be filled?*
- Who created the dataset (e.g., which team, partners and end-users)? *Reminder: is necessary to provide the associated grant (grantor and number of grant)*

5.1.2 Composition

Starting from the fact that the dataset will be leverage on open-source data, is necessary to pay attention to the composition. Beside to the fact that a dataset must be documented, this phase is important also to try to take control on bias that an unbalanced dataset could have on the AI training phase.

As the previous subsection, here the list of the key questions related to the composition is presented below:

- What do the instances that comprise the dataset represent (e.g., documents, photos, people…)?
- Description of the characteristic of each the instance types (data format, sampling frequencies, constraints as synchronisation…etc.)
- Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)?
- How many instances are there in total (of each type, if appropriate)
- **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** If it links to or relies on external resources, are there

guarantees that they will exist, and remain constant, over time; are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user?

5.1.3 Collection Process

This last subsection has the aim to put the base on the description of the collection process paying attention to the procedure both on the technical and on the ethical aspect.

Here the fundamental items for the collection process are listed below:

- **What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?** How were these mechanisms or procedures validated?
- **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**
- **Who was involved in the data collection process (e.g., project members participant, etc…)?**
- **Over what timeframe was the data collected?**
- **How was the data associated with each instance acquired?** Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data?
- **Did you collect the data directly, or obtain it via third parties or other sources (e.g., websites)?**
- **People are involved in the acquisition process? How many** (Knowing that is preferably that people involved are partners of the project)
- **Were any ethical review processes conducted (e.g., by an institutional review board)?** If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

5.2 Mapping of Data Collection VS Scenario Functionalities

This last part has the aim to summarise the collection process through a table. The last column will be filled in a second part of the project when the data will be effectively used.

Scenario: short description	Type of data & quantity	Scenario: functionality covered	Task (train/test/validation)

6 Conclusions

The main idea of this work is to create a dataset useful for First Responders application. To reach this purpose, we would like to take advantage of existing datasets commonly used for other kind of applications, typically home assistant, video surveillance, autonomous navigation, etc. The datasets presented will be reuse selecting the appropriate categories to comply with the end-users' requirements in the framework of TeamAware project. The data selected in literature belong to 3 three main applications

1. Human pose estimation and human activities dataset to reuse for Victim detection tasks
2. Situational awareness using visual segmentation. A particular focus will be put on the Smoke/fire detection
3. Damage assessment datasets

The dataset creation task is very ambitious. Moreover, knowing the complexity of the operational framework in which LEAs operate, WP3 and WP4 partners agreed to use other sources of data to complete this task in order to guarantee the performances and quality of the results of the AI algorithms. Indeed, synthetic data and ad-hoc data collection in the project test bed are forecasted.

The next version of the data set will be include images delivered by end users and collected in the test beds sites. For this purpose is planned to send to the end- a short story telling focalising the scene and the characteristic of the data necessary to complete the datasets.

In order to finalise the dataset is also planned to create to create a data aggregation methodology to merge all this resources together.

To publish a well-defined dataset covering the TeamAware objectives is to refine the methodology illustrated in chapter 5. The methodology proposed will ensure to publish an exhaustive datasheet enabling further applications in First Responders domain.

7 References

[1] S. H. Bach, B. He, A. Ratner and C. Re, “'Learning the structure' of generative models without labeled data,” ICML, pp. 273–282, 2017.

[2] M. Andriluka, L. Pishchulin and P. B. Gehler, “2D Human Pose Estimation: New Benchmark and State of the Art Analysis,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2014.

[3] T. Li, J. Liu, W. Zhang, Y. Ni, W. Wang and Z. Li, “UAV-Human: A Large Benchmark for Human Behavior Understanding with unmanned aerial vehicles,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 16266--16275, 2021.

[4] M. J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. a. C. I. Woodbridge, R. McConville, R. Santos-Rodriguez, K. Chetty and others, “OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors,” 2021.

[5] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio and T. Serre, “HMDB: a large video database for human motion recognition,” *IEEE 2011 International conference on computer vision,* pp. 2556--2563, 2011.

[6] S. Hadfield and R. Bowden, “Hollywood 3D: Recognizing Actions in 3D Natural Scenes,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 3398--3405, 2013.

[7] J. Wang, Z. Liu, Y. Wu and J. Yuan, “Mining actionlet ensemble for action recognition with depth cameras,” *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1290--1297, 2012.

[8] “Switzerland Aerial Drone footage dataset,” https://www.kaggle.com/kmader/drone-videos.

[9] “USTC Forest smoke and fire dataset,” http://smoke.ustc.edu.cn/datasets.htm.

[10] “Smoke & Fire,” https://www.kaggle.com/kutaykutlu/forest-fire.

[11] S. d. d. set, https://cvgl.stanford.edu/projects/uav_data/.

[12] “ Stanford dronedaset compressed,” https://www.kaggle.com/aryashah2k/stanford-drone-dataset .

[13] “Aerial data set,” https://www.kaggle.com/jangsienicajzkowy/afo-aerial-dataset-of-floating-objects.

[14] “Git LADI data set,” https://github.com/ladi-dataset.

[15] “LADI Registry,” https://registry.opendata.aws/ladi/ .

[16] “LADI data,” https://ladi.s3.amazonaws.com/index.html.

[17] “Indoor data,” http://buildingparser.stanford.edu/dataset.html.

[18] D. C. Luvizon, D. Picard and H. Tabia, “2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning,” CVPR, 2018.

[19] W. L. e. a. M. J. Bocus, “OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors,” 2021.