TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

D4.2.1 v1

Framework and Guidelines for the development of applications

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Supported by

AUTHORS LIST

Álvaro García Martín (agm) alvaro.garcia@uam.es

HISTORY

Version	Date	Editor	Description
0.0	03/03/2017	Álvaro García Martín	Initial draft version
0.1	17/034/2017	Álvaro García Martín	First draft version
0.2	21/03/2017	José M. Martínez	Comments
0.3	23/03/2017	Álvaro García Martín	Final Working Draft
1.0	25/03/2017	José M. Martínez	Editorial checking

CONTENTS:

1. INTRODUCTION 1

1.1. MOTIVATION 1
1.2. DOCUMENT STRUCTURE 1

2. FRAMEWORK AND GUIDELINES FOR THE DEVELOPMENT OF APPLICATIONS 3

2.1. FRAMEWORK FOR THE DEVELOPMENT OF APPLICATIONS 3
2.1.1. Hardware and computation environment 3
2.1.2. Development framework 3
2.2. GUIDELINES FOR THE DEVELOPMENT OF APPLICATIONS 4

3. CONCLUSIONS AND FUTURE WORK 7

APPENDIX A: QUESTIONNAIRE FOR APPLICATIONS OF HIGH AVAILABILITY VIDEO ANALYSIS. I

1. Introduction

1.1. Motivation

This deliverable describes the work related with the task T.4.2: Use Cases and Demonstrators. The task T.4.2 in work package 4 (WP4) aims at evaluating and integrating the algorithms developed within WP1, WP2 and WP3, in order to conform the global analysis chain to provide solutions to long-term video analysis for people behaviour understanding.

Based on the algorithms developed within WP1, WP2 and WP3, we aim to develop several demonstrators: For these reason, the objective of this task T.4.2 is the development of applications of the algorithms developed in the project, providing applications both for developers, *devapp*, (for evaluation and testing) as well as for final users, *demonstrators*, (use cases to be defined with the help of the Observing Partners). The use cases will focus on the applications areas of the surveillance and people monitoring challenges targeted by the project (e.g., outdoor surveillance, people monitoring in malls, in-home monitoring).

The applications will validate that the developed algorithms and approaches can be application enablers for helping in real cases for automatic surveillance, in-home monitoring (independent living), automatic shop monitoring, automatic patients monitoring in residences, etc. The demonstrators may be transferred to industry to further development or for launching technology transfer projects using the project results as background technology. The following companies have explicitly shown their commitment to act as Observing partners monitoring the project results in order to evaluate possible collaborations and technology transfer: CuevaValiente, Grupo Sanyres, NaevaTec, Tecnalia, and Treelogic.

In order to develop the applications with the help of the Observing Partners, we designed a methodological analysis of use cases of interest. A questionnaire was designed in order to identify interest areas that was submitted to Observing partners and other companies. Unfortunately, the response was low (only 2-3 returned questionnaires), and therefore the methodological analysis of results for defining use cases was not possible. All the questionnaires are available in Appendix A.

1.2. Document structure

This document is composed of the following chapters:

- Chapter 1: Introduction to this document.

- Chapter 2: Framework and Guidelines for the development of applications.
- Chapter 3: Conclusions and future work.
- Appendix A: Questionnaire for applications of high availability video analysis.

2. Framework and Guidelines for the development of applications

In this document, we present the existing framework and guidelines for the development of applications and demonstrator in relation with the project.

2.1. Framework for the development of applications

2.1.1. Hardware and computation environment

The Video Processing and Understanding Lab has currently a complete infrastructure for the development of the Project, including computing facilities, both for personal computing as well as for massive image processing, an image processing framework, an image acquisition infrastructure and several video datasets for performance evaluation.

The image processing infrastructure is composed by 5 Rack-mounted servers with high processing capacity, for running multiple concurrent virtual machines for algorithm development and testing. The Image acquisition infrastructure is composed by a set of fixed cameras (3 x SONY DFW-X710 cameras and 3 x SONY SNC-RZ50P PTZ cameras) installed in different locations of the EPS-UAM main building, several portable cameras (Pionner piA 1600, Basler 1900, Sony Exview, …), several webcams and alternative depth and infrared cameras (KINECT V1 and KINECT V2) for allowing flexible acquisition configurations.

The software infrastructure is based in both Unix and Windows operating systems (OS) and the main programming language is C++ with the OpenCv libraries.

2.1.2. Development framework

Since there is a great range of possible applications with different requirements and outcomes, the proposed framework for the development of applications has been designed in order to be efficient, scalable and as general as possible. We identify three main factors: video input format, video analysis module and final outcomes or results format.

According to the specific application (*devapp* or *demonstrator*), the video input format requirements include, at least, the options of one or multiple online cameras and or previously recorded video sequences, either for evaluation purposes or for training and configuration. In addition, the application could include additional inputs or initialization information or scene context.

The video analysis module must be written in C++ language, but other languages could be called as native functions from C++, for example: Matlab, python, etc.

Finally, according to the specific application, the results must be directly visualized through the visual interface and stored, both visual results and analytic ones (e.g., bounding boxes, trajectories, event detections) for further evaluation purposes (*devapps*) or data analysis (*demonstrators*). The latter may be part of the demonstrators, like for example an additional graph showing the number of people present in the scene during a period of time, a map of occupancy of an area.

2.2. Guidelines for the development of applications

In order to visualize the results or outcomes of the applications or demonstrators, they should include a visual interface. Each application includes at least two versions of the visual interface: the developer version (*devapp*) and one or several client versions (*demonstrators*). Both version should allow the selection of the input video format between a previously recorded video or an online camera.

The developer version is defined in order to help the possible user to define or evaluate visually the performance of the application in a new scenario. For this reason, the developer version includes not only the visual input and output of the application but also different intermediate results, configuration parameters and possible additional information about the application behaviour or performance. Figure 1 shows one example of a developer interface layout.

The client version is defined in order to simply visualize the application or demonstrator final outcome. For this reason, the client version is a much simpler interface version which includes optionally the visual input and at least the online result for each specific application, for example: trajectories, alarms, people density estimation, etc. Figure 2 shows one example of a client interface layout. Additionally, the demonstrator may show some data analytics.

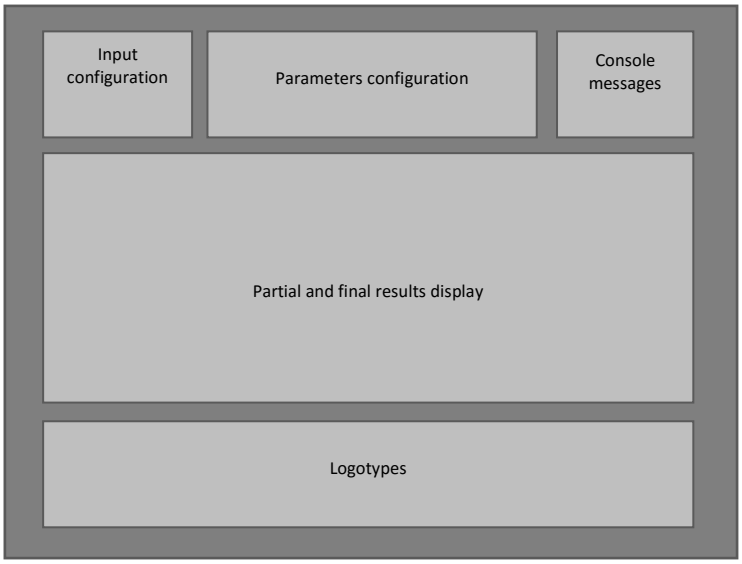

Figure 1. **Developer version general layout example**

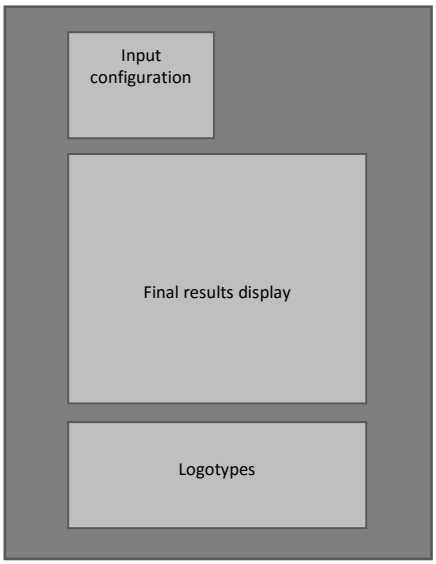

Figure 2. **Client version general layout example**

3. Conclusions and future work

In this document, we describe the task T.4.2. The objective of this task is the development of applications of the algorithms developed in the project, providing applications both for developers, devapp, (for evaluation and testing) as well as for final users, demonstrators, (use cases to be defined with the help of the Observing Partners). We present the existing framework and guidelines for the development of applications and demonstrator in relation with the project. In addition, in this document is also described the questionnaire designed in order to identify interest areas that was submitted to Observing partners and other companies.

Appendix A: Questionnaire for applications of high availability video analysis.

In order to develop the applications with the help of the Observing Partners, we designed a methodological analysis of use cases of interest. A questionnaire was designed in order to identify interest areas that was submitted to Observing partners and other companies. Unfortunately, the response was low (only 2 returned questionnaires), and therefore the methodological analysis of results for defining use cases was not possible. All the questionnaires are available in this Appendix A.

TEC2014-53176-R High Availability Video Analysis for People Behaviour Understanding
http://www-vpu.eps.uam.es/HAVideo/

QUESTIONNARIE FOR APPLICATIONS OF HIGH AVAILABILITY VIDEO ANALYSIS

We invite you to take part in a survey being conducted by the HAVIDEO project regarding your interests on the analysis tools for human behavior understanding and final applications for the task "T.4.2 Use Cases and Demonstrators". As Observing partners and final users monitoring the project results, we consider your opinion as a pivotal element in a successful completion of the project.

An initial preselection of use cases is already provided focusing on applications for people monitoring challenges targeted by the project (e.g, outdoor surveillance, people behavior in malls, in-home monitoring).

We thank you for your valuable input

ALL INFORMATION WILL BE TREATED IN THE STRICTEST CONFIDENCE

Only the researchers involved in the project will have access to the information in this survey. All data will be stored in compliance with University Autónoma of Madrid guidelines.

About the HAVIDEO project:

The HAVIDEO project addresses the high availability video analysis, that is, the long-term analysis of video sequences. It investigates strategies to model the monitored scene, detect the existing people and identify their behavior (e.g., movements and interactions) over long periods of time, either in single or multiple camera setups. The project results may underpin a wide range of applications ranging from security (i.e., video-surveillance) to health care (e.g., elderly home monitoring).

The key research elements of this project are the exploitation of the scene context, a continuous self-performance analysis, long-term adaptation and the use of multiple and possibly also complementary, sensors.

The HAVIDEO considers the development of demonstrators during the third year as a key element to validate the project achievements in the task "T.4.2 Use Cases and Demonstrators". These demonstrators will provide applications both for developers (e.g. evaluation and testing) as well as for final users (use cases to be defined with the help of the Observing Partners).

Additional information at http://www-vpu.eps.uam.es/HAVideo/

None

Video Processing and Understanding Lab

HA video

UAM UNIVERSIDAD AUTONOMA DE MADRID

TEC2014-53176-R High Availability Video Analysis for People Behaviour Understanding
http://www-vpu.eps.uam.es/HAVideo/

ANALISIS TOOLS FOR HUMAN BEHAVIOR UNDERSTANDING

Please indicate your level of interest in the tools listed in this page using a numerical scale from 1 (None) to 4 (High). Please use the boxes on the right side of this page.

DESCRIPTION	INTEREST
Segmentation Tools	
Hot initialization of Background subtraction algorithms Estimation of the scene background to initialize algorithms under challenging conditions such as crowds, shadows, global/local illumination and camera jitter.	High
Estimation of online performance/reliability for segmentation Identify whether algorithms are operating as expected (i.e. correctly segmenting objects) or a failure has occurred during runtime without ground-truth data.	Medium
Foreground segregation by fast-updating background modelling Especially suited for extremely crowded scenarios and for scenes recorded by a moving or shaking camera. This update is based on expected and observed data.	High
Detection tools	
People/object detectors New trained People/object detectors for custom applications including a wheelchair, wheelchair user and a parked car model.	High
Context-aware part-based people detector Local context is explored to account for non-visible parts of people due to scale constraints and for occlusions due to scene objects.	Medium
A Multi-configuration Part-based Person Detector A generic DPM multiple body parts combination framework in order to deal with specific partial occlusions in crowded scenarios.	Medium
Video tracking tools	
Multiple visual cues for particle filter based tracking Combination of multiple cues to improve robustness against tracking issues such as high speeds, occlusions, camouflages and objects with features similar to the target.	High
Estimation of online performance/reliability for video tracking Identify whether tracking algorithms are operating as expected (i.e. tracking the correct target) or a failure has occurred during runtime without ground-truth data.	Medium
Silhouette-wise object tracking Temporal propagation of tight-to-object masks by combining region segmentation and feature detection with tracking strategies under affine-transformation constraints.	Medium
Long-term object tracking with lost target recovery Single-object long-term tracker that supports target high appearance changes, occlusions, and is also capable of recovering a target lost during the tracking process.	High
Multi-target tracking with overlapped camera views Tracking a number targets not known a prior where targets may be observed simultaneously by multiple cameras	Medium
Miscelanea	
Resource modeling for camera networks Methods to estimate the consumption of resources related to the operation of cameras (i.e. energy consumption, communication bandwidth, processing load).	Low
Distributed framework for video analysis Infrastructure for distributed video analytics with descriptors for available tools which can be selected based on user/resource constraints (i.e. self-configurable).	Low
Inhibition of out-of-surface information in point-of-interest descriptions Points-of -interest are described only with information from surrounding surfaces that resemble the one of the point. The strategy succeed in finding inter-image correspondences when objects are surrounded by different backgrounds	Medium
Software for full simulation of multi-camera systems. This software may simulate the visual aspect (i.e. computer graphics), the camera hardware (i.e. their capabilities and resources) and the network interconnectivity.	Low

SELECTED APPLICATIONS FOR HUMAN BEHAVIOR UNDERSTANDING

Please indicate your level of interest in the applications listed in this page using a numerical scale from 1 (None) to 4 (High). Please use the boxes on the right side of this page.

DESCRIPTION	INTEREST
Abandoned object detection in long-term video-surveillance Detection of abandoned objects in public and (possibly) crowded places over long periods of time where scene conditions are likely to change (e.g. overall illuminations, number of people…)	High
Human behavior recognition for indoor monitoring Real-time recognition of simultaneous human activities (e.g. walking, raise hand) and interactions with the scene objects (e.g. leave, get & use) based on spatio-temporal features.	High
Automatic video editing software for lectures or classes A system to automatically generate an edited video and multimedia content using video object tracking algorithms and contextual information, such as the location of blackboard or slide presentation.	High
Online contextual updating in multi-camera scenarios Maintain and online-update a contextual description of a crowded scene. Identify the grade (none, punctual, continuous) of use of the contextual objects of the scene (doors, service desks, corridors, halls,…) and its status in each of the camera views (visible, occluded).	High
People counting based on multi-camera detection Combination of people detection algorithms along different cameras in order to estimate the number of persons in the scene.	High
Calibration of multi-camera scenarios GUI application to set, adjust or modify the main parameters of a multi camera scenario, such as number of cameras, multi-camera extrinsic calibration, common ground plane definition, etc.	Low

ADDITIONAL FEEDBACK

Please provide additional comments regarding the analysis tools and applications potentially interesting to be addressed within the scope of the HAVIDEO project.

Analysis tools for human behaviour understanding

Do you have any comments on the selected analysis tools in this survey? Would you like the HAVIDEO researchers to consider other tools related with the long-term analysis of video sequences? If yes, please state which ones.

Applications tools for human behaviour understanding

Do you have any comments on the selected applications tools in this survey? Would you like the HAVIDEO researchers to consider other applications related with the long-term analysis of video sequences? If yes, please state which ones.

Thank you again for your collaboration in this survey
The HAVIDEO project team

TEC2014-53176-R High Availability Video Analysis for People Behaviour Understanding
http://www-vpu.eps.uam.es/HAVideo/

QUESTIONNARIE FOR APPLICATIONS OF HIGH AVAILABILITY VIDEO ANALYSIS

We invite you to take part in a survey being conducted by the HAVIDEO project regarding your interests on the analysis tools for human behavior understanding and final applications for the task "T.4.2 Use Cases and Demonstrators". As Observing partners and final users monitoring the project results, we consider your opinion as a pivotal element in a successful completion of the project.

An initial preselection of use cases is already provided focusing on applications for people monitoring challenges targeted by the project (e.g, outdoor surveillance, people behavior in malls, in-home monitoring).

We thank you for your valuable input

ALL INFORMATION WILL BE TREATED IN THE STRICTEST CONFIDENCE

Only the researchers involved in the project will have access to the information in this survey. All data will be stored in compliance with University Autónoma of Madrid guidelines.

About the HAVIDEO project:

The HAVIDEO project addresses the high availability video analysis, that is, the long-term analysis of video sequences. It investigates strategies to model the monitored scene, detect the existing people and identify their behavior (e.g., movements and interactions) over long periods of time, either in single or multiple camera setups. The project results may underpin a wide range of applications ranging from security (i.e., video-surveillance) to health care (e.g., elderly home monitoring).

The key research elements of this project are the exploitation of the scene context, a continuous self-performance analysis, long-term adaptation and the use of multiple and possibly also complementary, sensors.

The HAVIDEO considers the development of demonstrators during the third year as a key element to validate the project achievements in the task "T.4.2 Use Cases and Demonstrators". These demonstrators will provide applications both for developers (e.g. evaluation and testing) as well as for final users (use cases to be defined with the help of the Observing Partners).

Additional information at http://www-vpu.eps.uam.es/HAVideo/

None

Video Processing and Understanding Lab

HA video

UAM UNIVERSIDAD AUTONOMA DE MADRID

TEC2014-53176-R High Availability Video Analysis for People Behaviour Understanding
http://www-vpu.eps.uam.es/HAVideo/

ANALISIS TOOLS FOR HUMAN BEHAVIOR UNDERSTANDING

Please indicate your level of interest in the tools listed in this page using a numerical scale from 1 (None) to 4 (High). Please use the boxes on the right side of this page.

DESCRIPTION	INTEREST
Segmentation Tools	
Hot initialization of Background subtraction algorithms Estimation of the scene background to initialize algorithms under challenging conditions such as crowds, shadows, global/local illumination and camera jitter.	High
Estimation of online performance/reliability for segmentation Identify whether algorithms are operating as expected (i.e. correctly segmenting objects) or a failure has occurred during runtime without ground-truth data.	Medium
Foreground segregation by fast-updating background modelling Especially suited for extremely crowded scenarios and for scenes recorded by a moving or shaking camera. This update is based on expected and observed data.	High
Detection tools	
People/object detectors New trained People/object detectors for custom applications including a wheelchair, wheelchair user and a parked car model.	High
Context-aware part-based people detector Local context is explored to account for non-visible parts of people due to scale constraints and for occlusions due to scene objects.	Medium
A Multi-configuration Part-based Person Detector A generic DPM multiple body parts combination framework in order to deal with specific partial occlusions in crowded scenarios.	Medium
Video tracking tools	
Multiple visual cues for particle filter based tracking Combination of multiple cues to improve robustness against tracking issues such as high speeds, occlusions, camouflages and objects with features similar to the target.	High
Estimation of online performance/reliability for video tracking Identify whether tracking algorithms are operating as expected (i.e. tracking the correct target) or a failure has occurred during runtime without ground-truth data.	Medium
Silhouette-wise object tracking Temporal propagation of tight-to-object masks by combining region segmentation and feature detection with tracking strategies under affine-transformation constraints.	Medium
Long-term object tracking with lost target recovery Single-object long-term tracker that supports target high appearance changes, occlusions, and is also capable of recovering a target lost during the tracking process.	High
Multi-target tracking with overlapped camera views Tracking a number targets not known a prior where targets may be observed simultaneously by multiple cameras	Medium
Miscelanea	
Resource modeling for camera networks Methods to estimate the consumption of resources related to the operation of cameras (i.e. energy consumption, communication bandwidth, processing load).	Low
Distributed framework for video analysis Infrastructure for distributed video analytics with descriptors for available tools which can be selected based on user/resource constraints (i.e. self-configurable).	Low
Inhibition of out-of-surface information in point-of-interest descriptions Points-of -interest are described only with information from surrounding surfaces that resemble the one of the point. The strategy succeed in finding inter-image correspondences when objects are surrounded by different backgrounds	Medium
Software for full simulation of multi-camera systems. This software may simulate the visual aspect (i.e. computer graphics), the camera hardware (i.e. their capabilities and resources) and the network interconnectivity.	Low

SELECTED APPLICATIONS FOR HUMAN BEHAVIOR UNDERSTANDING

Please indicate your level of interest in the applications listed in this page using a numerical scale from 1 (None) to 4 (High). Please use the boxes on the right side of this page.

DESCRIPTION	INTEREST
Abandoned object detection in long-term video-surveillance Detection of abandoned objects in public and (possibly) crowded places over long periods of time where scene conditions are likely to change (e.g. overall illuminations, number of people…)	High
Human behavior recognition for indoor monitoring Real-time recognition of simultaneous human activities (e.g. walking, raise hand) and interactions with the scene objects (e.g. leave, get & use) based on spatio-temporal features.	High
Automatic video editing software for lectures or classes A system to automatically generate an edited video and multimedia content using video object tracking algorithms and contextual information, such as the location of blackboard or slide presentation.	High
Online contextual updating in multi-camera scenarios Maintain and online-update a contextual description of a crowded scene. Identify the grade (none, punctual, continuous) of use of the contextual objects of the scene (doors, service desks, corridors, halls,…) and its status in each of the camera views (visible, occluded).	High
People counting based on multi-camera detection Combination of people detection algorithms along different cameras in order to estimate the number of persons in the scene.	High
Calibration of multi-camera scenarios GUI application to set, adjust or modify the main parameters of a multi camera scenario, such as number of cameras, multi-camera extrinsic calibration, common ground plane definition, etc.	Low

ADDITIONAL FEEDBACK

Please provide additional comments regarding the analysis tools and applications potentially interesting to be addressed within the scope of the HAVIDEO project.

Analysis tools for human behaviour understanding

Do you have any comments on the selected analysis tools in this survey? Would you like the HAVIDEO researchers to consider other tools related with the long-term analysis of video sequences? If yes, please state which ones.

Applications tools for human behaviour understanding

Do you have any comments on the selected applications tools in this survey? Would you like the HAVIDEO researchers to consider other applications related with the long-term analysis of video sequences? If yes, please state which ones.

Thank you again for your collaboration in this survey
The HAVIDEO project team