SEVENTH FRAMEWORK PROGRAMME

VIT
Vision for Innovative Transport

Project partly funded by the EC
Grant agreement no. *222199*

SP4-Capacities - Research for SMEs

REPORT ON PROTOTYPE VIDEO-SURVEILLANCE SYSTEM

Deliverable D5.2

Release date 15 January 2010

Work package number WP5
Work package title People security with video analysis
Activity Type RTD

About the Document

This document reports technical details of the final version of software package for robust object classification and tracking; it describes the content of the developed software modules, the documentation about how to use it, and the hardware specifications defined after field test period. Also it reports the activity of *Tasks 5.2 Statistical learning to support decision* and *Task 5.3 Video analysis and robust tracking* from months 13 to month 15 and the activity of *Task 5.4 Prototype integration and testing* from months 13 to month 18 within the Workpackage *WP5: People security with video analysis.*

The document has been produced by the collaboration of the workpackage WP5, the participants of the workpackage have all duly contributed to the activity of the workpackage and the production of this document and they endorse the final version as the conclusion of the workpackage.

Workpackage leader

Alberto Lovato (IMA)

Document authors

Francesca Odone (DISI)
Alberto Lovato (IMA)
Emanuele Trucco (DUN)

Document reviewers

Guido Porta (ILOG)
Thomas Keese (WITT)

Table of contents

INTRODUCTION

The purpose of the document is to report the last six months RTD activity within work package WP5. During these months RTD performers have selected the methods to be engineered in the final software release, and have tested them on the field in order to evaluate and to measure the obtained results.

We will report the hardware specification defined during the test period, the software pipeline used and the results obtained, we will present how the results respond to user requirements and finally we will give a documentation about how to use the software package.

Structure of the report

The report is organized as follows:

1. Overview
2. Testing scenario and hardware specifications
3. Software prototype
4. Experimental results
5. How software responds to user requirements
6. User manual

AUDIENCE

The present deliverable is filed as Confidential, as it contains critical information for the VIT project and also for the Metrocargo system.

Therefore the audience of the document is restricted the project participants --- the SME's who will find the technical details following their user requirements and the RTD performers who will use the present report as a guideline of their research and development activity.

1. Overview

WP5 aims to study and develop video analysis methods which are able to locate people in off-limit areas during load/unload operation of the Metrocargo system. Software analyses video cameras video flow and sends alert messages to the system when people are detected.

The main objectives of the WP are formalised in tasks T5.2 and T5.3. Task T5.4 refers to the integration and field test of the prototype. In the following we relate the main objectives of the tasks with the sections of this report.

T5.2 Statistical learning to support decision

Implementation of state-of-the-art statistical learning classification methods to discriminate humans from other moving objects in the scene.
Months 13-15
See Section 3.2 *"Classification methods to support decisions"*

T5.3 Video analysis and robust tracking

Design and implementation of high level video analysis functions, including tracking of moving objects and event abstraction. Tracking from multiple cameras will also be evaluated.
Months 13-15
See Section 3.1 *"Video analysis and robust tracking"*

T5.4 Prototype integration and testing

Integration of the software and hardware designed solution in the Metrocargo prototype for testing on the field.
Months 13-18
See Section 4 *"Experimental results"*

2. Testing scenario and hardware specifications

The field test has been carried out in the Vado Ligure plant where a prototype of the Metrocargo system is installed and operative. The area is around a stretch of railway of 60 meters and the prototype covers an area of about 800 m^2. Load/unload operations are made on one side of the railway where the containers are also stocked; in this sub-area happen most of mechanical devices actions.

The sub-area is covered by 4 TVCC cameras installed by the project leader ILOG to monitor the Metrocargo plant with the help of remote surveillance of a private institute; the cameras are colour Day&Night, 320 TV lines and are equipped with an infrared in order to better see during night time.

In the next figures it can be seen four pictures of the cameras.

Figure 1 – A view from video-surveillance Camera 1

Figure 2 – A view from video-surveillance Camera 2

Figure 3 – A view from video-surveillance Camera 3

Figure 4 – A view from video-surveillance Camera 4

The position of these cameras have been defined to monitor the widest possible part of the plant; moreover the purposes of the remote monitor of the plant carried out by human operators did not require high quality cameras. More specifically, these cameras were not intended for automatic video-surveillance, and people appears to be quite small making the video analysis hard.

RTDs have decided to enlarge the video-surveillance system with two more cameras for the specific needs of WP5 analysis. To this purpose the WP Leader IMAVIS (IMA) has installed the following cameras:

- One 3 **Megapixel** IP digital **Camera**, sensor CMOS 1/2,5", sensitivity 0,5 Lux F 1.2, MPEG4 and MJPEG compression, lenti C/CS Mount 1/2", Ethernet port 10/100/1000
- One Speed Dome **PTZ Camera** Day-Night, Wide Dynamic Range, CCD ¼ Sony progressive scan, 480 TV lines resolution, optical zoom 23X 3.6~82.8mm, digital zoom 12X, 0.1 Lux min. in colour mode, 0.02 Lux min. in BW mode.

The MegaPixel digital camera has been installed because these kind of cameras cover very wide areas without loosing resolution and at the same time provide digital PTZ functionalities which allow to view the details of a selected area without stopping the view of the complete frame. Megapixel cameras are the reference model for the near future of the video-surveillance market and their costs will rapidly decrease while their market share will grow. For these two main reasons RTDs are very confident that Megapixel technology is the best choice for the purpose of the WP5.

The PTZ camera has been installed in order to change remotely the viewing area and the zoom options. Several possible fields of view have been tested to better understand how the video analysis works with different framing and different size of the moving objects. This helped RTDs to simulate the behaviour of Megapixel cameras (considering both wide and detailed views) without having to manually move the installed Megapixel camera.

In the following two pictures we show a 3 megapixel full view of the camera (8 times the PAL resolution) and a detail of the same picture cut in the centre of it at PAL resolution.

Figure 5 – 3 megapixel full view from Megapixel Camera

Figure 6 – A detail of the previous view from Megapixel Camera at PAL resolution

Also, WP Leader IMA has installed ICSVision Analyser server on the Vado Ligure plant in order to:

- record videos from the different cameras installed;
- run and test the software both with live sequences and recorded videos.

ICSVision Analyser server is equipped with CPU Pentium IV 3.0Ghz, RAM 512Mb, Ethernet 10/100 Mbit, Hard Disk 250 Gb, 4 channel frame grabber for video PAL acquisition and digitalization, Linux OS, kernel 2.6.23.9, Debian etch distribution and all the needed software libraries.

Scenario summary

Vado Ligure Metrocargo plant is covered by 6 video cameras which monitor all the sensible area, including off-limit zones where the mechanical devices work and adjacent areas where unauthorised actions may start, and a server for recording, live view and video analysis:

- four cameras have been installed for remote video-surveillance purpose: they are PAL colour cameras, Day&Night with a 320 TV lines low resolution. Two of them monitor the off-limit zone while the other two watch the adjacent areas.
- the Megapixel digital camera, 3 megapixel resolution has been installed to monitor a wider area without loosing detail levels and it is pointed on the off-limit zone.
- the PTZ PAL camera, 480 TV lines high resolution, 23X optical zoom has been installed to simulate and test different framing and to add one camera on the off-limit zone.
- The ICSVision Analyser server has been installed to gather videos form the cameras, to record them continuously or on motion detection, to let the operator live viewing of the connected cameras and to analyse the video flow coming from cameras; the server has 4 input channel to manage video PAL and also manage the IP megapixel camera connected via Ethernet.

Summarising, 5 cameras are connected with the ICSVision Analyser server:

- o n. 4 video-surveillance PAL cameras (two of them pointed at the off-limit zone);
- o n. 1 Megapixel IP camera (pointed at the off-limit zone).

PTZ PAL camera has been used for testing different possible framing.

3. Software prototype

The software prototype receives as an input a video stream from each camera.

Besides a software initialization module including background initialization and parameters tuning, the prototype is made of two modules.

The first software module *(object detection and tracking)* follows this pipeline for each frame:

- Background update;
- Change detection and blob computation;
- Blob matching based on Kalman tracking;
- Kalman filter update.

In a given time instant the output is the list of objects detected on the scene for each camera – to each object is associated its history from its first appearance to the current instant.

The second software module (*object classification*) is applied to each detected blob to the purpose of checking whether it is associated to a human presence; the solution implemented on the final release is based on the analysis of the blob contour regularity along a time frame, according to the following pipeline:

- Set the reference blob (for instance the 5th occurrence of a blob on a tracked object)
- Compute the shape distance between the current blob and the reference one (from the 6th occurrence on);
- Updating the distance running variance;
- Classify the object based on the above distance and a predefined threshold;
- Send an alert after 3 consecutive coherent positive decisions.

The output is the alert message when a person is detected on the monitored area.

We now give a more detailed account of the implemented modules.

3.1 Video analysis and robust tracking

Given a video-camera under analysis, the object tracking module receives as an input a video stream and performs:

- **background update**, based on analysing the gray-level (or brightness) variance of each pixel in a fixed temporal interval and updating only those pixels with a small variance;
- **change detection**, using a background subtraction technique to detect the image regions moving with respect to the background and then building the connected components of pixels in motion, thus identifying moving blobs or units;
- **object tracking** on the moving blobs using an algorithm based on the position of the centroid of each blob which compensates segmentation errors, noise or occlusions, through a dynamic filter based on Kalman.

For the object tracking it is used an efficient data association method allowing for a correct tracking in situations where a group of moving objects overlap.

As a special case, sudden stops of objects have been modelled. The temporal interval of the background update process can be configured to regulate the elapsed time before an object enters in the background, thus disappearing from the list of moving objects. Within this time, the algorithm keeps on detecting the stopped object, also assisted by tracking that expects to find the object near to the predicted position.

The whole processing pipeline performs real-time on each camera video stream.

In this RTD activity, interactions between cameras have been studied following this two guidelines:

- **multi-camera tracking**: in order to make tracking more robust (limiting the number of ambiguities and tracking objects for longer periods) it is possible to combine tracking information for two adjacent cameras; the second camera can be initialised with the tracking information of the first camera, being faster and more precise in tracking the moving object. Tests carried out have shown that tracking multi camera is robust if the overlap between the two adjacent cameras is at least a quarter of the viewed frame; currently, no pairs of installed cameras meet this requirement, and Megapixel field of view may be set to cover the entire area of interest.
- **“adjacent area” vs “off-limit zone” cameras**: cameras can be divided in two categories, “off-limit zone” cameras which monitor areas with mechanical devices and “adjacent area” cameras which monitor neighbourhoods. Object detection and tracking module may be applied to the video input coming from the “adjacent area” cameras as a in-system pre-alert for “off-limit zone” camera, helping the vision system to be faster and more confident in the critical areas (in particular limiting the effects of occlusions). A batch set of tests on video-sequences acquired on the field has been carried out showing the usefulness of the approach. For a full operative implementation the access to the Metrocargo control system would be needed, for this reason the module is scheduled for future integration with Metrocargo.

3.2 Classification methods to support decisions

During the first 12 months three methods have been investigated, as detailed in Deliverable D5.1. In the last six months of the project, “shape regularity description” has been chosen as the method for the final version of the software.

This method has confirmed fast computation and very good results during field tests; moreover it does not show any performance degradation changing the test environment (Metrocargo system vs car/people) and the type of cameras, thus the more computationally expensive approach of the *size functions* is not necessary.

Here we briefly recall the method, referring to Deliverable D5.1 for more details.

The idea behind the algorithm is that the shape of a moving mechanic object is more stable and constant than the shape of a person walking in the scene.

The algorithm analyses the shape of the tracked object along time. For each tracked blob the algorithm sets a reference shape after a certain number of observations (5 observations in the final implementation), when we can assume the object is not a false alarm (i.e., a false positive). Then the algorithm compares the shape of each view with the shape of the reference one, returning a shape distance D.

Then the algorithm updates a running variance of the distance. At each frame the classifier decides if the object is mechanic object or not on the basis of a threshold: below the threshold the object is mechanic, above is a person. The method robustness is improved by the exploitation of temporal continuity: before an alert is given a certain number *N* of consecutive coherent answers have to be associated to each moving objects. A meaningful *N* is estimated on a selected training set and has been set to 3.

4. Experimental results

The field test carried out in the Vado Ligure plant has been divided in two parts: an offline test part during month 16 and a live test part during 17 and 18 months.

4.1 Offline test (month 16)

The offline part of the test has been based on selected recorded videos.

The ICSVision Analyser server has been configured in order to continuously record video shots of all the cameras with a motion detection function to save interesting (motion) events; then all the events has been analysed and a set of 150 videos has been selected and manually labelled.

The videos have been chosen to include different environment conditions (cloud, sun, rain, variable weather) and different day times. The length of the videos are from 1 to 5 minutes, with an average length of 3.4 minutes; the frame rate of the videos is 16 frame per second, with an average number of 3.264 frames each.

The videos have been collected from the camera available at month 16: the four video-surveillance cameras and the Megapixel camera.

50 videos form the training set and they have been used for training and validating the software modules and to find the right parameter tuning. The remaining 100 videos have been used for testing.

4.2 Live test (months 17-18)

The live part of the test has been run on the ICSVision server installed in the site.

A final version of the software has been installed and configured on the server to get the video live sequences from the cameras and to analyse them in real time. This final version has been trained and validated using all the 150 videos and the parameters of the offline test part.

Each moving object is tracked and classified (as described in the previous section 3.) and the software sends an event with the type (humans or mechanical device) of the object and correspondingly draws the contour of the object on the output video with a different colour for the two possible types.

The server has been configured to continuously record the cameras and to save in the database all the events where an object has been classified as human or mechanical device. Also, the motion detector module has been run and all the related events has been saved in the database.

The live test has been run for two months and has been qualitative evaluated as follows:

- videos recorded in the presence of “classified object” events have been verified to check false positive;
- videos recorded in the presence of motion events have been verified to check false negative.

During this part of the test, PTZ camera has been installed (end of month 16) to be used for validating results coming from megapixel camera without the need of physically moving it.

4.3 Results

The field tests aim to give an evaluation about the capability of the software to classify humans against mechanical man-made autonomously guided devices in motion. The software has been analysed as a single module getting a video as input and giving back an event when a moving object has been classified as a human.

On the 100 recorded and manually labelled videos we have the following error percentage:

- 0.7 % false positive;
- 3,6 % false negative.

A qualitative evaluation of the results obtained on live tests confirmed these results. The size of the analysed events is very satisfactory mainly for two reasons:

- researchers testing the device for WP3 field test were frequently around them, creating a typical danger situation
- continuous data acquisitions over long temporal spans highlighted similar situation of people moving around other type of moving mechanical devices as single locomotive or whole trains

We noticed a few exceptions in the obtained results that we are going to describe.

The use of the 4 video-surveillance cameras has pointed out the difficulty of using such cameras which were not installed for video analysis purposes. Moreover, cameras were not maintained and, especially in the rainy month 18, the camera case misted up increasing the error rate in the analysis.

RTDs decided to keep on using that cameras because they brought a big amount of data but they are definitely not the right cameras for the analysis purposes.

On the other side, Megapixel Camera has shown excellent robustness and stability with respect to the change of weather conditions. The only problematic situation occurred in the presence of very strong wind – the camera bracket is too loose for the severe wind exposure of the Vado Ligure plant.

The final view of the Megapixel camera chosen for the live test part has been defined to be fully compatible with the WP4 task of reconstructing train profile and reading container codes, as planned in the technical specifications. Results are particularly good also for this important aspect.

We conclude with a remark on night vision. As it can be seen in the pictures 7 and 8, the difference of the vision during the night between Megapixel e video surveillance camera is quite substantial. In the case of Megapixel the number of false alarms has grown up, but we are confident that the illumination of the plant, currently missing, will further improve the scene visibility and consequently reduce the number of false alarm back to standard conditions.

Figure 7 – A night view from video-surveillance Camera 4

Figure 8 – A night view from Megapixel Camera

In the next pages we show some pictures coming from video analysis: object classified as people are orange bounded while the other are blue bounded.

Figure 7 – People detected on the video-surveillance Camera 4

Figure 8 – Mechanical Device detected on the video-surveillance Camera 4

Figure 9 – Mechanical device detected on the video-surveillance Camera 3

Figure 10 – People detected on the video-surveillance Camera 3

Figure 11 – People and mechanical devices detected on the video-surveillance Camera 3

Figure 12 – People and mechanical devices detected on the video-surveillance Camera 3

Figure 13 – People detected on the video-surveillance Camera 2

Figure 14 – Mechanical device detected on the video-surveillance Camera 2

Figure 15 – People detected on the Megapixel Camera

Figure 16 – A detail of the previous view

Figure 17 – People detected on the Megapixel Camera

Figure 18 – A detail of the previous view

Figure 19 – Mechanical Device detected on the Megapixel Camera

Figure 20 – A detail of the previous view

5. How software responds to user requirements

We are going to summarise user requirements, constraint and measurable objectives and how the software responds to them.

How software responds to user requirements

- **Detect the presence of people in the work area**
  - The software segments the moving objects through a change detection module with a background updating model based on the variance of each pixel in grayscale
  - The software tracks the moving objects (blob) with the help of the Kalman filter, taking into account the stopped object thanks to a slow background updating
  - The software classifies every blob using a shape regularity description; a configurable number of consecutive frame with a coherent classification makes the method robust.
- **Keep under control the whole terminal to monitor and automatically report any operational irregularity and safety breach, in the first place regarding human safety**
  - The software runs simultaneously on different cameras covering the critical areas of the terminal and automatically alerts the system in case of human presence detection where it is not allowed for safety reasons.
- **Issue a signal than can be used to stop all equipment in the area where the human presence was detected (safety) and to alert the control room (security)**
  - When human presence is detected in the off limit area, the algorithm sends an immediate alert to the system management in order to immediately stop the operations of load/unload and to alert security human operators.
- **Let the human operator monitor the area through video cameras**
  - Analysis server running on the terminal has an interface which allows the operator to continuously monitor the area through live and recorded video stream of the cameras.

How software meets constraints

- **Work in extreme weather conditions: fog, rain, snow, wind up to 100 km/h, temperatures between -30° and +50°C.**
  - Hardware component installed in the terminal meet these constraints and let the software analyses videos without specific problems. Special attention has to be put in the choice of case and bracket to limit stabilisation problems due to strong wind.
  - Image processing methods to contrast noise due to bad weather has been implemented and successfully tested. A strong fog has been defined as the most critical situation.
- **Operate in daylight and at night, preferably using its own illumination although artificial lighting can be provided**
  - All the cameras are day-night video-cameras and are sufficiently illuminated for the purpose. Tests have been successfully carried on.

How software meets measurable objectives

Following the above mentioned motivations, considering that people intrusion in the plant may be seen as a very rare event (considering the complexity of the anti-intrusion system an estimate of 1 intrusion per month, mainly due to errors of human operators is an upper bound) the following error percentages are acceptable:

- **Maximum error percentage -- people detection in the work area:**
  - **in standard conditions: false alarms 5%, misses 1%**
  - **in extreme conditions: false alarms 8%, misses 3%**
  - **at night: false alarms 8%, misses 3%.**

The tests above described meet the objectives defined by the SME in the User Requirements.

6. User manual

The user manual of the final version of the software has no differences with respect to the months 12 version. We report the whole section for the convenience of the reader.

The software prototype is an executable compiled for Linux OS with the following requirements:

- server equipped with CPU Pentium IV 3.0Ghz or higher, RAM 512Mb min, Ethernet 10/100 Mbit, Hard Disk 80 Gb min;
- frame grabber for video PAL acquisition and digitalization;
- Linux OS, kernel 2.6.23.9, Debian etch distribution
- low level libraries distributed with the software

WP Leader IMA has placed its ICSVision Analyzer server in order to run and test the software.

The program takes in input and analyzes a standard YUV420P video stream captured from a CCTV camera.

The configuration of the program is done by an **XML file** like the following:

```
<peopledetection>
  <background interval="100" alpha="0.075" threshold="60" />
  <tracker maxDist="200" min-area="200" maxColorDist=”0.3” />
  <regions>
    <rectangle name="work area">
      <point y="296" x="72" />
      <point y="332" x="416" />
    </rectangle>
    <background width="640" height="480" />
  </regions>
  <shape-analysis min-observation=”5” variance-threshold="0.3" alarm-
        threshold=”3” />
  <rpcclient listener="10.0.0.23:9000"/>
</peopledetection>
```

The background node controls the behaviour of the change detection module used by the algorithm, here follows a brief description of each parameter:

- *interval="100"* the number of frames an object should be static to be included in the background
- *alpha="0.075"* the weight given to new frames in the background updating
- *threshold="60"* threshold used in the difference between current frame and the background to detect moving objects

The tracker node controls the behaviour of the object tracker used by the algorithm, here follows a brief description of each parameter:

- *maxDist="200"* maximum distance (in pixel) allowed for an object in two consecutive frames to be followed by the tracker
- *min-area="200"* minimum area of objects to track
- *maxColorDist="0.3"* maximum distance between colour histograms of objects to track

The regions node in the XML controls the regions of interest in the scene. In this way it is possible to select only some regions of the scene to control for the presence of people.

- *Rectangle points* set the vertexes of rectangles where analysis has to be performed
- *BG width and height* the coordinates of the points are calculated with respect to this size of the frame