A multi-camera image processing and visualization system for train safety assessment

Giuseppe Lisanti · Svebor Karaman · Daniele Pezzatini · Alberto Del Bimbo

Abstract In this paper we present a machine vision system to efficiently monitor, analyse and present visual data acquired from a railway overhead gantry equipped with multiple cameras. This solution aims to improve the safety of daily life railway transportation in a two-fold manner: (1) by estimating multiple safety requirements using image analysis algorithms that can process large imagery of trains (2) by helping train safety operators to detect any possible malfunction on a train. The system exploits high-rate visible and thermal cameras that observe a train passing under a railway overhead gantry. The machine vision system is composed of three principal modules: (1) an automatic wagon identification system, recognizing the wagon ID according to the UIC classification of railway coaches; (2) a system for the detection and localization of the pantograph of the train; (3) a temperature monitoring system. These three machine vision modules process batch trains sequences and their resulting analysis are presented to an operator using a multitouch user interface.

Keywords System · Machine vision · Train · Safety
1 Introduction

In the last years, several disastrous train accidents have brought train safety to the media’s and public’s attention, as those that happened in Italy in 2009 [21] and in France in 2013 [3]. Train accidents may be a result of either a problem on the railway tracks, as it was the case for the French accident, or some issues with the train itself. The analysis of the railway tracks requires the installation of sensors on board of a train that should travel on the tracks that have to be inspected. Several proposals have been made in this sense [7, 13]. This work focuses on the safety assessment of the train itself.

A train is composed of a locomotive and multiple wagons, any of its components can be a risk for the train safety. A single wagon failure can trigger the derailment of several wagons and have dramatic consequences. The Viareggio accident [21] is believed to be the consequences of an axle failure on a tank wagon, the wagon hit the platform of the station and overturned to the left and several following wagons also overturned, exploded and caught fire. Another important aspect of train safety is temperature monitoring especially when the train is approaching a tunnel where escape in case of fire can be difficult. For example, the Kaprun disaster [30] was due to an electric fan heater that caught fire. Hence monitoring an abnormal temperature on any part of the train can provide an early notice of an issue and thus prevent its potential dramatic outcome. A train can be considered safe for transit if all wagons are adapted for the transit on the railway, the locomotive and all wagons exhibit nominal temperatures and no out of shape elements are present. Failure to fulfill any of these requirements may indicate a risk situation. The analysis of each train status may be done by stopping and analyzing each train in a offtrack location before being allowed to travel. This would induce serious delays on the train traffic. Hence, more interest have been put on portal based system that could be installed on some important keypoints of the railway network (e.g. before a tunnel or before entering a train station) in order to assess on-the-fly all the safety requirements. This solution has the clear advantage of not requiring to stop the train to run the analysis and it is also possible to install multiple sensors on a single portal providing a thorough analysis of the train status at once. However, such portal based approaches require the monitoring system to be able to capture all required signals even for a train running at full speed.

This paper depicts our proposed multi-camera portal for train safety assessment. The system relies on high-speed visible and thermal cameras to monitor several aspects of the train. The acquired signals are processed by computer vision methods to extract meaningful information. Finally, all the information is provided to an operator through a touch-based user interface. We first review in the next section the state-of-the-art of computer vision based system for train safety and of touch-based interaction for control rooms and train safety. We then give an in-depth presentation of our proposed system in Section 3, specifying the hardware and giving an overview of the software developed to obtain reliable data acquisition from all sensors. In Section 4, we detail how we solve each task of the train analysis, namely the automatic wagon identification, the detection and localization of the pantograph of the train, and the temperature monitoring. We then describe how all the results are provided to the operator on our multitouch interface. Finally in Section 5, we give an evaluation of each sub-system of our multi-camera portal on a dataset recorded in an ad-hoc railway test-bed.
2 Related work

2.1 Computer vision based systems for train safety

We can distinguish in the literature the approaches that target safety assessment of the railway or the train itself. Some approaches focus on a single aspect of train safety, while multi-modal portals try to analyse at the same time multiple safety characteristics.

Many railway accidents happen at railway crossing where an object such as a car is stopped on the railway, hence one common use of computer vision is to detect if an obstacle is obstructing the railway. Machine vision was used in [28] to detect moving obstacle in these locations. A 3D vision system for obstacle detection is proposed in [35]. Recently, authors of [33] have addressed the problem of railway visual inspection by proposing a computer vision technique based on superpixels. Experimental results show that this railway recognition method outperforms other approaches existing in the literature and could be used to perform automatic anomaly detection. Train stations are also a risk environment, in [12] the authors presented a method for automatically detecting people jumping or falling off a train platform.

Another aspect to consider to assess train safety is the proper configuration of the train itself. In [29], a system to detect misalignment of a train pantograph was proposed. The authors of [16] proposed a multi-function portal similar in spirit to ours with the main objectives of detecting misalignment of carriage or abnormal temperature so as to be able to stop a train before it enters a tunnel. The system exploits line-scan cameras to obtain the train image in the visible domain, pyroelectric line cameras for thermal imaging and a distributed time-of-flight telemeter for the train shape analysis. The evaluation targets mostly the sensors performance and only qualitative results of out-of-shape detections are given. Another portal based system for train safety assessment was proposed in [1], where a calibration procedure for cameras placed on a train inspection portal is presented. The calibration of the sensors allows the reconstruction of the entire 3-D train profile and the creation of a thermal map of the wagons.

Finally, several existing railway safety system include methods for automatic wagon identification. This information is especially essential when the acquired data are analysed by an operator. A recent survey [19] has reviewed the commercially available computer vision systems of wagon identification. This task is related to the more general text detection and recognition problem in computer vision. The work from Matas et al. [23] proposing the Maximally Stable Extremal Regions (MSER) has been highly influential for the text detection problem, since text characters are usually detected as MSER. This method has been shown [24] to be a very effective and stable descriptor for affine regions. Its sensitivity to blur has been improved years later in [9], by combining it with the Canny edge detection method [8]. Text recognition, dubbed OCR for Optical Character Recognition, is often addressed using open source libraries such as Tesseract\(^1\) or relying on commercial products like OmniPage or ABBYY Fine Reader. It has also been shown in [17] that dealing with

\(^{1}\)https://github.com/tesseract-ocr/tesseract
the specificity of a document or a task can improve the recognition. However, text detection and recognition is still an open research problem as suggested by the active ICDAR challenges [18].

2.2 Touch-based interface for control rooms and train safety

Operators in control rooms are often asked to monitor multiple safety characteristics and have to perform crucial security operations in a short amount of time. For these reasons, information visualization and touch-based interaction play a key role when developing a monitoring tool for a control room.

Several studies [15, 20] have been done to assess the benefits of the adoption of a multitouch workstation for tasks that require interaction with multiple visual cues. Results have shown that multitouch interaction can be twice as fast as a mouse based one. Furthermore, multitouch interaction is often preferred by users since it offers direct manipulation of graphical elements, resulting in a more natural and effective approach to carry out the requested tasks. Touch-based interaction has been exploited in control and security processes since the early seventies. In 1973, Beck and Stumpe [4] proposed a prototype of touchscreen device to control the new CERN accelerator. In recent years, studies were conducted to propose and evaluate good practices in the design process of touch-based interfaces for security operators. In particular, Zahler [37] proposed multiple patterns for the design of touch-based user interface for railways security and other safety-critical applications. In [5], the authors investigated the effectiveness of direct manipulation in multitouch interfaces for safety-critical situations in maritime control room. Results showed that direct manipulation of interface elements can enhance the situational awareness of users.

The evaluation and testing of safety-critical interfaces is crucial to show whether a novel developed system actually fulfills its goals. Authors of [32] proposed a method for the evaluation of an user interface for safety in railway based on a high-fidelity simulator of an interlocking systems. Although many standardized usability evaluation methods exist and are commonly used for general purpose systems, some specific methods have been defined for the evaluation of safety critical interactive systems [34].

2.3 Contributions

Our proposal is to use high-speed cameras mounted on a railway overhead gantry to monitor multiple aspects of the train. To this end, we designed and developed a hardware/software solution that can simultaneously manage several high-speed sensors with different speeds. We then perform multiple image analysis of the acquired data in order to assess the train safety. In particular, the proposed solutions aim at:

1. automatically segment the wagon identifier according to the UIC classification of railway coaches;
2. detect the pantograph passage;
3. extract the wagon temperature to allow anomaly detection.

Finally, we propose a touch-based interface that adopts common interaction metaphors like direct manipulation and multitouch gestures [2]. The goal of the touch-based user interface is to give operators a quick and efficient way to interact with the results of the image analysis.
3 Our multi-sensors portal

We designed and developed a hardware/software solution to manage multiple sensors, and combine different types of media (visual images, video sequences, thermal images). In the following, we describe the physical structure and the sensors characteristics of our multi-camera portal. We then detail the acquisition manager we designed to manage all the sensors together and deal with their high-rate acquisition.

3.1 Architecture of the portal and sensors involved

The portal is built over a single rail, has a height of 8.5 meters and is 6 meters large. The distance from the train side is about 1.5 meters. This gantry is equipped with a total of 5 sensors, the Fig. 1 illustrate our portal configuration. We used three different types of high rate sensors. We summarize the characteristics of each sensor in Table 1. Two of these sensors operate in the visual spectrum (one linear and the other matricial) while the last one works in the thermal spectrum (with linear acquisition).

In particular, the matrix camera operating in the visual spectrum (Teledyne Dalsa HM-640) acquires 300 grayscale images per second at a resolution of 640x480 pixels. This

Fig. 1 The portal equipped with all cameras. The visual matrix camera (HM-640) is on top in the center of the gantry, two thermal cameras (256L) are positioned on each side of the portal, and two visual line cameras (Spyder 4K) are positioned on the same side but one on top of the other
camera is positioned on the top and at the center of the portal and is used as a general overview of the train passing by.

The linear camera operating in the visual spectrum (Teledyne Dalsa Spyder 4K) acquires 18500 grayscale lines per second with a height of 4096 pixels. We have positioned two linear cameras on the side of the portal: one is used to observe the bottom and central part of the train and is the entry signal of the wagon identification system, see Section 4.1; the second linear camera is positioned higher on the side of the portal to capture the top of the train and it will be used for the detection of the pantograph detailed in Section 4.2.

Finally, the linear camera operating in the thermal spectrum (PYROLINE 256L) acquires 512 lines per second of 256 pixels each with a temperature range of $[30^\circ \ldots 800^\circ]$°. One of this thermal sensor is positioned on the top of each side of the portal, and they are used to monitor the temperature anywhere in the passing train as explained in Section 4.3.

### 3.2 Acquisition framework

We designed a specific hardware/software solution that allows us to obtain a good synchronization between all the sensors, enabling a meaningful and easily interpretable playback of the acquisition for the operator.

As regards hardware, three separate servers are used to deal with the large amount of data induced by the high-rate cameras mounted on the portal. Specifically, there are 2 servers acquiring data from one linear and one thermal camera, while the third server deals only with the matricial camera as shown in the overview of our system architecture given in Fig. 2. These servers have 8 SAS disk in RAID-01 to obtain the sufficient speed needed (about 270MB/s) to write all the data generated, while maintaining sufficient reliability.

Concerning the software, an effective solution have been implemented that allows us to simultaneously control each sensor focusing on the optimization of CPU, memory and disk usage. In particular, we designed a frame-grabber for each type of sensor: visual matrix, visual linear and thermal linear.

As shown in Fig. 2 each frame grabber is controlled by a HTTP server (Acquisition-Server). For each camera this server implements some acquisition primitives (e.g. start, stop, pause) as well as some specific function depending on the camera model, for example, focus control for the thermal cameras.

To simultaneously control each AcquisitionServer (of each sensor) we designed another HTTP server called the AcquisitionManager. Each AcquisitionServer registers to the AcquisitionManager and periodically send its state. Once an AcquisitionServer is registered to the AcquisitionManager it is possible, using a simple web interface we developed, to control the IP address, the state of the grabber, and all the primitives expected for the relative sensor.

The acquisition primitives in common between all sensors are shown as a unique button in the web interface of the AcquisitionManager, while the primitives specific to a sensor are shown only for the registered sensor that can use them. Having a common interface showing the state of all sensors is particularly useful, for example, to prevent starting a

| Camera model          | Data type | Type  | Hz   | Resolution  | Data rate  | # of Sensors |
|-----------------------|-----------|-------|------|-------------|------------|--------------|
| Teledyne Dalsa HM-640 | Visual    | Matrix| 300  | 640x480     | 92MB/s     | 1            |
| Teledyne Dalsa Spyder 4K | Visual | Line  | 18500| 4096x1      | 80MB/s     | 2            |
| PYROLINE 256L          | Thermal   | Line  | 512  | 256x1       | 128KB/s    | 2            |
new acquisition or stopping an ongoing acquisition if some of the AcquisitionServers are still saving some recently acquired data. When an AcquisitionServer is closed it unregisters from the AcquisitionManager. This software design offers the advantage that a sensor can be easily added, activated or deactivated for a specific acquisition.

Full synchronization of all the sensors can be challenging due to the high and different rates of acquisition. Our acquisition framework have been designed so that each AcquisitionServer runs on a different core of a server and communicate with a single camera through a dedicated GigE network interface. In order to avoid clock drift, each AcquisitionServer clock is synchronized to the AcquisitionManager clock using the Network Time protocol. When at time $t$ a primitive is called from the AcquisitionManager, it sends the request with a start time $t + \delta$ to all the AcquisitionServers, where $\delta$ is a sub-second delay used to guarantee start synchronization. Our hardware setting ensures timely acquisition. Finally, given all the acquisitions, data from one sensor can be synchronized with the other sensors’ data based on their respective acquisition rate.

Although this solution has proved to work effectively, each camera in our system have also a general purpose input/output (GPIO) connector that could be used to receive or control external signals (e.g. start recording on train passage). Therefore, even stronger synchronization could be obtained using dedicated hardware.

### 4 Train analysis

The different types of media acquired with the high-rate cameras are processed exploiting state-of-the-art techniques, in order to extract relevant information and assess the train safety requirements. In particular, we developed three sub-systems. The first sub-system is responsible for the detection of the wagon identifier of each wagon. Due to the high dimension of the images acquired with the high-speed cameras (4096 × 18500 pixels per
second) a solution for fast and robust wagon id segmentation is proposed. It first performs basic blob segmentation and then uses a weighted voting scheme of RANSAC inliers to find the region related to the characters of the wagon identifiers. The second subsystem is responsible for the detection of the pantograph and its segmentation from the image of the whole train. For this task, we adopt a local features matching approach using the SIFT [22] features. Finally, the third sub-system is responsible for monitoring the temperature of the train.

4.1 Wagon identification

The Wagon identifier subsystem aims at identifying the wagon by segmenting its unique international identification number from the image acquired with the Visual Line Camera 1 positioned at the bottom right of the portal, see Fig. 1. From this identifier multiple characteristics can be extracted (type of wagon/locomotive, owner and country for example) and thus one can understand if the wagon is expected and allowed to transit on the monitored railway section. Due to the huge dimension of the image and the presence of noise we need to apply a robust identifier segmentation method. The whole method, described in Algorithm1, relies on image processing and geometric analysis to obtain the position of the identification number in the image.

**Algorithm 1** Wagon ID segmentation.

**Input:** $I, r_D, d, s, w, h$

**Output:** $b$

1. Compute $\tau_O = \text{AdaptiveThreshold}(I)$;
2. Extract edge $I_e = \text{EdgeDetector}(I, \tau_O)$;
3. Perform morphological dilation $I_D = (I_e \oplus \text{disk}(r_D))$;
4. Perform hole filling $I_f = \text{fill}(I_D)$;
5. Extract Connected Components $c_{bbox} = \text{extractCC}(I_f)$;
6. Initialize votes $v \leftarrow 0$;
7. Initialize $j \leftarrow 0$;
8. Initialize $k \leftarrow 0$;
9. while $j \leq w$ do
10. \hspace{1em} while $k \leq h$ do
11. \hspace{2em} $c_{bbox} = \text{SelectCC}(c_{bbox}, j, k, d)$;
12. \hspace{2em} $n = \text{FitLine}(c_{bbox})$;
13. \hspace{2em} $v = v + \text{Voting}(c_{bbox}, n)$;
14. \hspace{2em} $k = k + s$;
15. \hspace{1em} end
16. \hspace{1em} $j = j + s$;
17. end
18. $b = \text{SegmentSalientRegions}(I, c_{bbox}, v)$;

Given an image of a wagon $I$ of width $w$ and height $h$, we first apply an adaptive thresholding method to find the optimal threshold that separate foreground from background pixels, such as:

$$\tau_O = \text{AdaptiveThreshold}(I)$$  \hspace{1em} (1)

The Otsu adaptive thresholding algorithm [27] assumes that the image contains two classes of pixels (e.g. foreground and background) then calculates the optimum threshold separating these two classes in order to minimize intra-class variance. After that, the threshold $\tau_O$ is
used as input for the Canny edge detector [8] to segment the contour of the foreground elements present in the image.

\[ I_e = EdgeDetector(I, \tau_O). \]

(2)

The regions defined by connected edges are filled first by using the morphological operation of dilation and then with a fill operation to definitely close small holes:

\[ I_D = I_e \oplus disk(r_D), \]

(3)

\[ I_f = fill(I_D), \]

(4)

where \( r_D \) represent the disk ray size used for the dilation operation. Once those regions are filled, a connected components labelling algorithm is used to define the bounding boxes containing the blob regions previously segmented:

\[ c_{bbox} = extractCC(I_f). \]

(5)

In order to identify the connected components corresponding to characters, we first filter out regions that have an area value outside of a reasonable range, and then we apply a voting procedure based on the sliding-window paradigm. In particular, given all the bounding boxes of the connected components we can infer that the characters of the identifier are close to each other and mostly aligned along a line. For this purpose, we apply a sliding-window procedure (with sampling step of \( s \) pixels) to the image and for each sub-window, of dimension \( d \times d \) pixels, we fit a line through the RANSAC algorithm [14] considering only the bottom-right points of the bounding boxes present in that sub-windows:

\[ \hat{c}_{bbox} = SelectCC(c_{bbox}, j, k, d), \]

(6)

\[ \text{in} = FitLine(\hat{c}_{bbox}), \]

(7)

where \( j \) and \( k \) represent the coordinates of the sub-windows considered. In each sliding-window, the points \( \text{in} \) selected by RANSAC as inliers accumulate a vote:

\[ v = v + Voting(c_{bbox}, \text{in}). \]

(8)

At the end of this procedure the points with the most votes will represent the bounding box with a higher probability of containing a character. Finally, to obtain the identifier which is composed of 12 characters we selected the sub-region of the image \( \hat{b} \) containing the most voted and aligned foreground objects. The alignment is estimated by computing the distances on the x-axis \( D_x \) and y-axis \( D_y \) for the 20 most important foreground regions according to \( v \). Then we take the exponential of the negative of these distances and we weigh the votes previously obtained with those matrices separately.

\[ v_w = \exp(-D_x) * v + \exp(-D_y) * v. \]

(9)

All those regions with a weighted vote \( v_w \) greater than zero represent a character of the ID, see the example in Fig. 3. From this image, any Optical Character Recognition (OCR) method can be applied to obtain the identifier. This identifier segmentation step is necessary as wagon image have an average size of 4096 \( \times \) 80000 pixels and cannot be processed as is by an OCR method.
4.2 Pantograph detection

The Pantograph detection subsystem detects the passage of the pantograph in order to avoid false positive cuts in a laser based system\(^2\) that analyse the shape of each wagon. The pantograph detection is run on the data extracted from the Visual Line Camera 2 positioned at the middle right of the portal. The obtained segmented high resolution image can also be analysed by an operator to determine if there is any anomaly in the pantograph shape.

The proposed solution is composed by an offline phase and an online phase. Offline, we extract SIFT [22] keypoints from an image of the pantograph used as template \(T\), see Fig. 4:

\[
D_T = \text{ExtractSIFT}(T).
\]

The SIFT descriptors are representations of image regions highly discriminative and invariant to changes in brightness, scale and rotations. We store the extracted descriptors in a KD-Tree \([10, 11, 25]\) in order to speedup the matching process:

\[
K_T = \text{KDT ree}(D_T).
\]

Online, once the mosaic image of the train side is obtained, the proposed Algorithm 2 extracts SIFT keypoints from that image:

\[
D_c = \text{ExtractSIFT}(I_c),
\]

\(^2\)This system is composed of three infrared laser mounted on the portal. This is a proprietary solution and cannot be discussed in the scope of this paper.
The KD-Tree nearest neighbor search provides the identifier of the closest descriptors, however, due to the curse of dimensionality, descriptors neighbors in $\mathbb{R}^{128}$ could be not visually similar. For this reason a second filtering is introduced. The distance between the first and the second most similar descriptors is measured and the match are discarded according to $\frac{d_1}{d_2} \leq \tau_d$, where $\tau_d = 0.67$, as in [6]:

$$D_m = \text{MatchDescriptors}(T, D_c, \tau_d).$$ (13)

However, these matches are not guaranteed to be correct, this can occur for various reasons, for example repeated structures in the image or points with similar SIFT descriptors. For this reason a third validation step is performed by applying a geometric robust validation following a projective model transformation. This is obtained by exploiting the RANSAC algorithm [14] to fit a projective model and successively by applying a consistency check algorithm to the estimated homography in order to determine if the fitted model is correct:

$$[H, \text{in}] = \text{FitProjective}(D_m);$$

$$p_{bbox} = \text{CheckGeomConsistency}(H, \text{in}).$$ (15)

In this way it is possible to establish the presence of the pantograph and it is also possible to have an indication of its location in the image and segment the relative sub-image to be shown to the operator.

In real world scenarios, multiple types of trains with potentially different pantographs with different appearance could transit on the railway. The proposed KD-Tree indexing scheme would in this situation index all SIFT features extracted from all templates. Our first step of analysis, i.e. finding matching SIFT could still be performed to find a set of candidates pantograph templates. The geometric verification with RANSAC must then be performed for each candidate template.

### 4.3 Temperatures segmentation

The Thermal monitoring subsystem acquires two thermal maps of each wagon and compare them to nominal operating temperatures in order to issue an alarm in case of fire risk due to abnormally high temperatures. The sensors involved in this subsystem are the Thermal Line Camera 1 and 2 positioned at the top right and left of the portal, see Fig. 1.

Each thermal camera is connected and managed by a different server in order to ensure a higher robustness of the thermal subsystem through duplication. The mosaic image obtained from all the acquired lines concatenation is divided into subregions of fixed size and for
each subregion both the mean and maximum temperatures are calculated. The average and maximum temperatures coming from one camera are compared with those extracted from the other camera in order to validate their output and ensure that both the servers and sensors are working correctly.

An important phenomenon to be considered is the distortion of the temperature values caused by the perspective in the image. In particular, the pixels furthest from the center of the sensor will be subject to a high distortion caused by the perspective between the sensor and the observed wagon (obviously this phenomenon depends also on the distance from the sensor to the wagon). For this reason we used two cameras observing the wagon from two different viewpoints.

### 4.4 Touch-based user interface

To enable an operator to visualize and interact intuitively with the results of the wagon analysis we developed a touch-based graphical user interface (GUI) based on multitouch interactions. The aim of the GUI is twofold. On one hand, it is used to exhibit all the results of the acquisition and analysis to the security operator in a simple way so he can quickly get an overview of the train status. On the other hand, it provides the operator with several tools for a direct and easy manipulation of all the data necessary to assess the train safety requirements.

An operator first loads a session of a processed wagon analysis in the interface. A session is composed by (i) a frontal video of the train obtained from the matrix camera positioned at the top of our portal, (ii) two thermal images that are the output of the thermal monitoring subsystem, (iii) a high-resolution image of the train acquired by the visual linear camera. The frontal video can be played and scanned through a timeline visualizer. The timeline of the video is synchronized with visual markers on both the thermal and linear imagery in order to give a visual time reference on all results.

Thermal images are displayed with a false-color scale obtained from the temperature values. By default the scale is based on the $\text{max}$ and the $\text{min}$ values of the image, but the operator can manually change the range of colors in order to enhance the visualization of specific temperature values. Left and right thermal images can be activated with a selector, so that only one image at a time is visualised in the interface.

The linear camera acquisition result is a high-resolution image of the train. In order to allow a fluid and smooth manipulation of this image we adopted a multi-resolution tiling technique. Acquired images are pre-processed in order to have a set of downscaled versions of high-definition ones. Each downscaled version is then decomposed in tiles of 256x256 pixels. The rendering engine of the GUI loads and display only the tiles required for the current zoom level and portion of the image visualized by the operator, instead of loading the entire high-definition image. The operator can activate graphical overlays on the train image in order to visualize the results obtained by the pantograph detection subsystem and the wagon identifier subsystem.

Figure 5 shows an overview of the interface and some phases of the interaction of a control operator with the interactive GUI, like checking temperature of an area of the wagon or visualising a high-definition image of the train. The set of functionalities provided by the GUI allows the operator to have a quick overview of the image processing results and to perform punctual and precise controls through direct manipulation using multitouch gestures.
5 System evaluation

To evaluate our proposed system, we recorded a dataset of sequences using the portal depicted in Fig. 1 in Zmigrod, Poland. We acquired 34 sequences of a train composed of one locomotive and one wagon on a test-bed railway track of 1 Km. The train passed under the multi-camera gantry several times, at different speeds and with different weather conditions. To register these sequences we used the system architecture and the web interface previously described in Section 3. Each sub-system described in Section 4 has been evaluated quantitatively or qualitatively on the acquired dataset.

In particular:

- the wagon id segmentation algorithm has been evaluated on the images acquired with one of the two linear cameras and processed in order to assess the ability of our solution to segment each character and the whole identifier. We also compare our segmentation method with the MSER approach. Timings for the main steps of the adopted solution are also reported. More details are given in Section 5.1.
- The pantograph detection performance is reported in Section 5.2. Tests have been conducted using as template an image of the pantograph taken from a calibration sequence (in order to grant separation with the test sequences). We also performed a leave-one-out cross-validation in order to assess the generalization ability of the proposed solution on the acquired dataset. Timings for both features extraction and matching are also reported.
- Temperatures extraction from the two thermal cameras is discussed in Section 5.3.
- The user interface we proposed has been evaluated using an usability inspection method. More details are reported in Section 5.4.

Tests were conducted on an Intel Xeon@2.60GHz, and our method does not exploit parallel optimization or GPU hardware.

5.1 Wagon identification

As shown in Fig. 6, it would be really difficult to segment the identifier with a standard camera. Indeed, a high motion blur due to the train speed affects the readability of several...
characters. The use of a high-rate linear camera is hence necessary to be able to properly segment the wagon identifier. To evaluate the wagon identifier segmentation performance, we estimated the accuracy of both the full train ID and the single characters segmentation for each wagon. In particular, for the case of characters segmentation we count as true positive every detected region that contains a character of the wagon ID, as false negative every missed character of the wagon ID and as false positive every region classified as part of the ID but non containing a character of the wagon ID. While for the full ID segmentation, we count a true positive every time all the characters of the wagon ID are recognized, a false negative every time at least one character of the wagon ID is missed and as false positive all the regions classified as positive but that do not contain a character of the wagon ID. For the full ID segmentation evaluation, there is exactly one target detection by wagon making it easier to obtain higher false positive rate as any region that do not contains the ID will be counted as a false positive.

As it can be observed from Table 2 the accuracy of the system is very high for full ID segmentation in the case of wagon 1 while for the wagon 2 we are always able to detect the full train ID. We can also appreciate that both false positives and false negatives are limited for the full train ID segmentation of each wagon. When evaluating in terms of character segmentation, see Table 3, the results are even better with really low false negative and false positive rates.

Furthermore, we compare our approach to the MSER [23] method. Namely, we replace the steps 1 to 5 in our Algorithm 1 with MSER to obtain the candidate regions to be validated with our geometric verification step. The results given in Tables 2 and 3 show that our proposed method outperforms the MSER. We have observed that the MSER can fail to detect some characters on the second wagon, as also reported in Fig. 7, and this substantially lowers the results for the segmentation of the whole identifier.

In Fig. 8 we show a qualitative sample of how the proposed solution segments the train identifier, for both the locomotive and the wagon. One can observe how the train identifier is a very small part of the initial image and appreciate how our method successfully detects it.
### Table 3  ID characters segmentation accuracy

| Method | Accuracy | FN Rate | FP Rate |
|--------|----------|---------|---------|
| Wagon 1 | Ours     | 93.4    | 6.6     | 0.7     |
|        | MSER     | 85.1    | 14.9    | 1.2     |
| Wagon 2 | Ours     | 100.0   | 0       | 1.5     |
|        | MSER     | 71.6    | 28.4    | 14.0    |
| Average | Ours     | 96.7    | 3.3     | 1.1     |
|        | MSER     | 78.3    | 21.7    | 7.6     |

We also computed the timings needed to segment the identifier for each wagon. Timings for each salient step of our algorithm are reported in Table 4. These timings have been averaged overall the 34 sequences in our dataset. The most time consuming steps are the edge extraction (about 3 seconds) and the windowing with RANSAC (about 3 seconds). However, none of the steps of our algorithm have been optimized to run in parallel on multiple cores or exploit GPU hardware. The overall process computational cost is of about 7 seconds, as reported in the last line of Table 4.

### 5.2 Pantograph detection

The acquisitions were performed in two different settings, one of them with the pantograph outside the field of view of the visual linear camera. For this reason, the pantograph was observed for 18 (out of 34) sequences of the dataset. The pantograph template has been taken from a preliminary calibration sequence not included in the final dataset in order to ensure separation with the test data. For each one of the 18 sequences the pantograph was correctly detected by the proposed solution. In Fig. 9 we report some samples of the pantograph matching working under very different illumination and weather conditions.

Given the small set of available sequences we also performed a leave-one-out cross-validation, considering as template the pantograph of one sequence and as test all the remaining sequences, for all the sequences. In this setting, the detection accuracy was of 97.4 %, i.e. the detection only failed for 8 test sequences when the template of the pantograph was taken from a rainy and dark sequence.

We also computed the computational cost required for pantograph detection. The feature extraction time for the template image requires on average 0.12s, this operation is done one time when the system is initialized. Obviously the most time consuming operation is the extraction of the features from the high resolution image of the whole train (e.g. 4096 × 180000 pixels), which requires on average 4.3s. Finally, descriptors matching, RANSAC

---

**Fig. 7** Examples of false negatives and false positives when using MSER for blob segmentation
Fig. 8 Examples of text region segmentation on the two different wagons. In green the bounding box segmented after RANSAC refinement

fitting and the geometry consistency check requires just 0.03s on average. If a faster analysis would be required, the code could be further optimized to run in parallel on multiple cores or to exploit GPU hardware.

5.3 Temperatures segmentation

Temperatures have been acquired for all 34 sequences in our dataset from both thermal cameras. Performing a quantitative evaluation is not possible in this case mainly because during the session of tests it was not possible to measure the real temperatures of the train.

A qualitative example is shown in Fig. 10. Here we can appreciate the segmented train as observed from the two cameras. The hottest part (in red) is the locomotive engine.

As regards computational complexity, this subsystem meets the real-time requirements, since it just requires to compute average and maximum temperatures for subregions of fixed size, compare them between the two cameras and with a reference set of temperatures for alerting.

Table 4 Timings for each step of the wagon identification subsystem. All timings are in second and averaged over all the sequences of our dataset

| Step                                    | Wagon 1 | Wagon 2 | Average |
|-----------------------------------------|---------|---------|---------|
| Edge Detection                          | 2.56    | 3.50    | 3.03    |
| Dilate                                  | 0.30    | 0.37    | 0.33    |
| Fill                                    | 0.45    | 0.56    | 0.50    |
| Extract Connected Components            | 0.40    | 0.47    | 0.43    |
| Windowing, RANSAC and Voting            | 4.36    | 1.21    | 2.79    |
| Overall                                 | 8.07    | 6.11    | 7.09    |
5.4 Evaluation of the user interface

Cognitive Walkthrough (CW) [36] is a usability inspection method whose objective is to identify usability problems, focusing on how easy it is for new users to accomplish predefined tasks. It is a technique that aims at detecting errors in design that would interfere
Table 5  Report of usability inspection obtained with the Cognitive Walkthrough method

|      | Visibility errors | Feedback errors | Task completion (users) |
|------|-------------------|-----------------|-------------------------|
| T1   | 1                 | 0               | 5 out of 5              |
| T2   | 1                 | 2               | 5 out of 5              |
| T3   | 2                 | 0               | 4 out of 5              |
| T4   | 1                 | 1               | 5 out of 5              |
| Overall | 5                 | 3               | —                       |

with the performance of users while interacting with the tool. CW is usually carried out by specialists in the field of interface development and usability experts. As the walkthrough proceeds, comments of the users are recorded. Previous studies on usability testing [26] showed that the number of usability problems $U_p$ found in a usability test is:

$$U_p = N(1 - (1 - L)^n)$$

where $n$ is the number of users, $N$ is the total number of usability problems in the design and $L$ is the proportion of usability problems discovered while testing a single user. The typical value of $L$ is 31%, suggesting that 85% of usability issues can be found with 5 testers.

We conducted an usability inspection of the proposed touch-based interface to assess its possible usability issues. For this purpose we defined the following different tasks:

| Task | Description |
|------|-------------|
| T1   | Load the most recent analysed sequence |
| T2   | Position the video on the sequence corresponding to the pantograph detection |
| T3   | Visualize areas of the wagon which temperature is higher than 50° |
| T4   | Visualize the wagon ID number using the analysis results |

For each task, we defined a sequence of actions with details about specific task flow from beginning to end. We asked 5 users to perform the defined task using the so-called think aloud technique in order to record failure in the interaction and design suggestions. According to the CW test workflow proposed in [31], for every action we assess: 1) if the users know or discover what to do at the current step (visibility) and 2) if the users realize they have done the right action and are making progress towards their goal (feedback). When users highlight an issue concerning one of these two aspects, we record a visibility error or a feedback error.

Only one user was not able to complete task T3, while the others completed all the assigned tasks, reporting usability and user interface related problems while performing the evaluation. We report in Table 5 the rate of success for each task, along with the number of usability issues found during the execution of the tasks. The overall usability of the interface resulted adequate to complete the proposed tasks. Furthermore, we used feedback from users to identify and correct 8 usability issues, mostly regarding visibility, sizes and positions of objects on the screen or ambiguities in the use of textual labels.

6 Conclusions

In this paper we introduced a multi-camera portal for train safety assessment. Our proposal is able to perform the analysis of multiple safety requirements of a train passing under the gantry. We detailed the hardware used and the software developed to robustly acquire data...
from multiple high-rate sensors. Image processing and computer vision methods are applied
on each data stream to extract meaningful information. We also presented our multitouch
interface that enables an operator to quickly observe and simply interact with the processed
data. The evaluation has shown the good performance of the analysis and the usability of
the interface.

Acknowledgments This work was supported by the Integrated Intermodal System for Security and Sig-
naling on Rail (SISSI) project, funded by Regione Toscana (Italy) under the PAR FAS 2007-2013 program
(P.I.R. 1.1.B, Action 1.1). We also thank Andrew D. Bagdanov and Iacopo Masi for their support in the
project realization.

References

1. Andria G, Bruno A, Lanzolla AML, Spadavecchia M, Scarano VL (2014) Camera calibration proce-
dure to improve safety in railway tunnel. In: Proceedings of the IMEKO TC4 International Symposium
and the International Workshop on ADC Modelling and Testing Research on Electric and Electronic
Measurement for the Economic Uptum
2. Baraldi S, Del Bimbo A, Landucci L (2008) Natural interaction on tabletops. Multimedia Tools and
Applications 38(3):385–405. doi:10.1007/s11042-007-0195-7
3. BBC-News (2013) Deadly french train crash at bretigny-sur-orge. http://www.bbc.com/news/
world-europe-23294630
4. Beck F, Stumpe B (1973) Two devices for operator interaction in the central control of the new CERN
accelerator. CERN. Tech rep
5. Bjørneseth FB, Dunlop MD, Hornecker E (2012) Assessing the effectiveness of direct gesture
interaction for a safety critical maritime application. Int J Hum Comput Stud 70(10):729–745.
doi:10.1016/j.ijhcs.2012.06.001
6. Brown M, Lowe DG (2003) Recognising panoramas. In: Proceedings of the International Conference on
Computer Vision
7. Camargo LFM, Edwards JR, Barkan CP (2011) Emerging condition monitoring technologies for railway
track components and special trackwork. In: Proceedings of the Joint Rail Conference
8. Canny J (1986) A computational approach to edge detection. Trans Pattern Anal Mach Intell 8(6):679–
698. doi:10.1109/TPAMI.1986.4767851
9. Chen H, Tsai SS, Schroth G, Chen DM, Grzeszczuk R, Girod B (2011) Robust text detection in natu-
ral images with edge-enhanced maximally stable extremal regions. In: Proceedings of the International
Conference on Image Processing
10. Del Bimbo A, Lisanti G, Pernici F (2009) Scale invariant 3D multi-person tracking using a base set of
bundle adjusted visual landmarks. In: Proceedings of the International Conference on Computer Vision
Workshops
11. Del Bimbo A, Lisanti G, Masi I, Pernici F (2011) Continuous recovery for real time pan tilt zoom local-
ization and mapping. In: Proceedings of the International Conference on Advanced Video and Signal
Based Surveillance
12. Delgado B, Tahboub K, Delp EJ (2014) Automatic detection of abnormal human events on train
platforms. In: Proceedings of the National Conference on Aerospace and Electronics
13. Edwards JR (2009) Advancements in railroad track inspection using machine-vision technology. PhD
thesis, University of Illinois at Urbana-Champaign
14. Fischler MA, Bolles RC (1981) Random sample consensus: A paradigm for model fitting
with applications to image analysis and automated cartography. Commun ACM 24(6):381–395.
doi:10.1145/358669.358692
15. Forlines C, Wigdor D, Shen C, Balakrishnan R (2007) DiRect-touch vs. Mouse Input for Tabletop
Displays. In: Proceedings of the Conference on Human Factors in Computing Systems
16. Fumagalli L, Tomassini P, Zanatta M, Libretti G, Trebeschi M, Sansoni G, Docchio F (2012) Reliability
and safety in railway. In: Tech, chap Multifunction Portals for Train Monitoring: Recent Advances and
Innovative Optoelectronic Instrumentation
17. Kae A, Huang G, Doesch C, Learned-Miller E (2010) Improving state-of-the-art OCR through high-
precision document-specific modeling. In: Proceedings of the International Conference on Computer
Vision and Pattern Recognition
18. Karatzas D, Gomez-Bigorda L, Nicolaou A, Ghosh S, Bagdanov A, Iwamura M, Matas J, Neumann L, Chandrasekhar VR, Lu S et al (2015). In: International Conference on Document Analysis and Recognition

19. Kazanskiy N, Popov S (2015) Integrated design technology for computer vision systems in railway transportation. Pattern Recognit Image Anal 25(2):215–219. doi:10.1134/S1054661815020133

20. Kin K, Agrawala M, DeRose T (2009) Determining the benefits of direct-touch, bimanual input on a multitouch workstation. In: Proceedings of the Conference on Graphics Interface

21. Landucci G, Tognoli A, Busini V, Derudi M, Rota R, Cozzani V (2011) The viareggio LPG accident: Lessons learnt. J Loss Prev Process Ind 24(4):466–476. doi:10.1016/j.jlp.2011.04.001

22. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110. doi:10.1023/B:VISI.0000029664.99615.94

23. Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. Image Vis Comput 22(10):761–767. doi:10.1016/j.imavis.2004.02.006

24. Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Van Gool L (2005) A comparison of affine region detectors. Vision Int J Comput Vis 65(1-2):43–72. doi:10.1007/s11263-005-3848-x

25. Muja M, Lowe DG (2009) Fast approximate nearest neighbors with automatic algorithm configuration. In: Proceedings of the International Conference on Computer Vision Theory and Applications

26. Nielsen J, Landauer TK (1993) A mathematical model of the finding of usability problems. In: Proceedings of the Conference on Human Factors in Computing Systems

27. Otsu N (1979) A Threshold Selection Method from Gray-level Histograms. Trans Syst Man Cybern 9(1):62–66. doi:10.1109/TSMC.1979.4310076

28. Pu YR, Chen LW, Lee SH (2014) Study of moving obstacle detection at railway crossing by machine vision. Inf Technol J 13(16):2611–2618. doi:10.3923/itj.2014.2611.2618

29. Sacchi M, Cagnoni S, Spagnoletti D, Ascari L, Zunino G, Piazzini A (2011) PAVISYS: A computer vision system for the inspection of locomotive pantographs. In: Proceedings of the International Conference on Pantograph Catenary Interaction Framework for Intelligent Control

30. Schupfer H (2001) Fire disaster in the tunnel of the kitzsteinhorn funicular in kaprun on 11 nov 2000. In: Proceedings of the International Conference on Safety in Road and Rail Tunnels

31. Spencer R (2000) The streamlined cognitive walkthrough method working around social constraints encountered in a software development company. In: Proceedings of the Conference on Human Factors in Computing Systems

32. Stelzer A, Schütz I, Oetting A (2014) Evaluating Novel User Interfaces in (Safety Critical) Railway Environments. In: Proceedings of the International Conference on Human-Computer Interaction. Applications and Services

33. Teng Z, Liu F, Zhang B (2016) Visual railway detection by superpixel based intracellular decisions. Multimedia Tools and Applications 75(5):2473–2486. doi:10.1007/s11042-015-2654-x

34. Thimbbleby H (2007) Interaction walkthrough: evaluation of safety critical interactive systems. In: Proceedings of the Workshop on Interactive Systems. Design, Specification, and Verification

35. Weichselbaum J, Zinner C, Gebauer O, Pree W (2013) Accurate 3D-vision-based obstacle detection for an autonomous train. Comput Ind 64(9):1209–1220. doi:10.1016/j.compind.2013.03.015

36. Wharton C, Rieman J, Lewis C, Polson P (1994) Usability inspection methods. Wiley, New York, pp 105–140. chap The Cognitive Walkthrough Method: A Practitioner’s Guide

37. Zahler T (2008) A design process for constructing a user interface pattern library for touch-based applications in safety-critical environments. In: Proceedings of the International Conference on System Safety
**Giuseppe Lisanti** received the PhD degree in computer science from the University of Florence, Italia. He is a postdoc at the Media Integration and Communication Center and his main research interests focus on computer vision and pattern recognition, specifically for person detection and tracking, person re-identification, 2D and 3D face recognition.

**Svebor Karaman** received the PhD degree in computer science from the University of Bordeaux, France. He is currently a postdoctoral researcher at Columbia University, NY, USA. His research interests include computer vision and machine learning, more specifically video analysis and indexing, action recognition, person re-identification, face recognition and hashing.
Daniele Pezzatini received the master degree in computer science from the University of Florence. He is currently a researcher at the Media Integration and Communication Center. His main research interests focus on interaction design, natural interaction systems and interactive web applications.

Alberto Del Bimbo is Full Professor of Computer Engineering, and Director of the Media Integration and Communication Center at the University of Florence. He was the Deputy Rector for Research and Innovation Transfer of the University of Florence from 2000 to 2006. His scientific interests are multimedia information retrieval, pattern recognition, image and video analysis and natural human-computer interaction. He has published over 300 publications in some of the most distinguished scientific journals and international conferences, and is the author of the monography “Visual Information Retrieval”. He is IAPR Fellow, associated editor of several leading journal in the area of pattern recognition and multimedia, and Editor-in-Chief of the ACM Transactions on Multimedia Computing, Communications, and Applications. He also received the prestigious SIGMM 2016 Award for Outstanding Technical Contributions to Multimedia Computing, Communications and Applications.