A Video-Based Object Detection System for Improving Safety at Level Crossings

N. Fakhfakh[*,1,2], L. Khoudour[1], E.M. El-Koursi[1], J. Jacot[2] and A. Dufaux[2]

[1]*French National Institute for Transport and Safety Research (INRETS): 20, rue Elisée Reclus 59666 Villeneuve d'Ascq, France*

[2]*Ecole PolytechniqueFédérale de Lausanne (EPFL): CH-1015 Lausanne, Switzerland*

Abstract: Improving transport users' safety is one of the main priorities of research into transport system attractiveness. Level crossings are one of the most critical weak point involving road and rail users' infrastructure. They have become increasingly dangerous and unsafe due to road and railway users' behavior. Furthermore, rail and highway safety professionals from several countries must deal with the same subject: providing safer level crossing. Actions are planned in order to exchange and share knowledge on existing level crossings technologies between academic organizations and industrial operators, and provide experiments for improving the management of level crossing safety and performance. This has enabled us to discuss sharing knowledge gained from research in order to improve safety at level crossings. This article provides research results about possible technological solutions to reduce the number of accidents at level crossings. As a main contribution is that we discuss and prove the effectiveness of the use of video sensing for object detection. Furthermore, we have tested and adapted a robust technique for moving object detection, which is followed by a new approach for 3D object localization.

Keywords: Image processing, level crossings, safety, technologies, vision.

1. INTRODUCTION

Within the past years, railways undertakings became interested in the assessment of level crossings safety. Level crossings have been identified as a particular weak point in road and railway infrastructure, seriously affecting their safety. Statistically, more than four hundred people year die in accidents involving road and vehicles at road-rail level crossings in the European Union (EU) [1]. 90% of these fatalities are linked to errors committed by road vehicle drivers. However, the behavior of road vehicle drivers and pedestrians cannot be previously estimated beforehand. It is important also to note the high cost related to each accident (approximately one hundred million euro per year in the EU for all level crossing accidents). For this purpose, road and highway safety professionals from several countries have dealt with the same subject: providing safer level crossings. Actions are planned in order to exchange information and provide experiments for improving the management of level crossing safety and performance. This has enabled us to discuss sharing knowledge gained from research into improving safety at level crossings. In recent times, the coordinated action for the sixth framework program entitled "Safer European Level Crossing Appraisal and Technology" (SELCAT) [2] has provided recommendations for further actions, evaluation of possible technological solutions to improve safety at level crossings. The proposed solution must reduce the risk presented by road drivers or railway

operators. Another interesting project was led by Japan entitled Intelligent Transport System (ITS) [3]. This project is steadily progressing and aims to make transportation systems safe and convenient. Furthermore, the evolution of technologies allows researchers to test and evaluate the performances of several devices at level crossings.

Nevertheless, the high safety integrity levels for technical solutions are defined by an approved common safety methodology. High safety requirements for level crossing systems mean a high cost which hinders the technological setup of advanced systems. The systems having unacceptable levels of risk must be eliminated and not be implemented. High technology systems are exploited and introduced in order to prevent collisions between trains and automobiles and to help reduce levels of risk from railroad crossings. Several conventional object detection systems have been tested on railroad crossings. These techniques provide more or less significant information accuracy.

An overview of existing technological solutions and a proposed object detection system based on a passive vision system, such as television cameras, constitute the main part of the research described in this paper. After an introduction covering the problem of level crossings safety and the motivations to improve safety at level crossings, the second part deals whit a survey of conventional obstacle detection system at level crossing. We discuss also in this part the advantages and the drawbacks of each system. The third part will be devoted to arguing for and describing the choice of the video sensing and image processing to alleviate the previous drawbacks. This is followed in the fourth section by outlining the proposed new system for object detection based on passive vision. An overview of the proposed system is

*Address correspondence to this author at the French National Institute for Transport and Safety Research (INRETS): 20, rue Elisée Reclus 59666 Villeneuve d'Ascq, France; Tel: +33 3 20 43 83 24; Fax: +33 3 20 43 83 98; E-mail: nizar.fakhfakh@inrets.fr

presented which will be divided into two parts: foreground detection based on a single camera and a novel approach for 3D localization of a detected moving foreground. The proposed approach is evaluated in the last part of this paper using conventional datasets of images and tested on real level crossing image dataset. The conclusion part is devoted to a discussion on the obtained results.

2. REQUIREMENTS AND CONVENTIONAL SENSORS AT RAILROAD CROSSING

2.1. Requirements

The most reliable solution to decrease the risk and accident rate at level crossings is to eliminate unsafe railroad crossings. This avoids any collisions between trains and road users. Unfortunately, this is impossible in most cases, due to location feasibility and cost that would be incurred. For instance, almost 10 million Euros per year are earmarked for the removal of the most dangerous level crossings in France. To overcome these limits, the development of a new obstacle detection system is required. Possible solutions using advanced technologies to improve level crossing safety will be discussed in this part. Any proposed system is not intended to replace the present equipment installed on each level crossing. The purpose of such a system is to provide additional information to the human operator, it can be considered as support system operations. This concerns the detection and tracking of any kind of objects, such as pedestrians, people on two-wheeled vehicle, wheelchairs and car drivers. Presently, sensors are evaluated relying on their false object detection alert among other. This may increase the risk related to level crossing users. It is important to be noted that risks associated with the use of technology systems are becoming increasingly important in our society. Risk involves notions of failure and consequences of failure. Therefore, it requires an assessment of dependability; this might be expressed, for example, as probability of failure upon demand, rate of occurrence of failures, probability of mission failure, and so on. Each level crossing is equipped with various sensors for tracking and timely detection of potentially hazardous situations. To be reliable, the related information must be shared and transmitted to the train dispatching center, stations, train drivers and road users. A brief summary of the technical characteristics of most devices provides the necessary understanding to design effective obstacle detection system. Generally, most level crossings are fitted with standard equipments such as lights, automatic full or half barriers, notices. This equipment warns and prevents all users of the level crossing if a train is approaching the dangerous area.

2.2. Conventional Sensors at Level Crossings (LC)

Today, there are a number of trigger technologies installed at level crossings, but they all serve the same purpose: they detect moving object when passing at particular points in the LC. Indeed, those conventional obstacle detection systems have been used to prevent collisions between trains and automobiles. Several technologies, such as optical beam provide this (Fig. **1**). The principle is as follows: optical emitters are placed on one part of the crossing, each one emitting a directed optical beam with a defined field of emission. Then a photon detector intersects the field of emission of the emitter. If the

beams are interrupted, it means that an object is located on the crossing.

Fig. (1). Optical beam method.

This technique has the advantage that is easy to replace, but it has many important drawbacks: very expensive cost for installation, need to have several detectors along the crossing, traffic needs to be stopped for installation, and it is unusable in period of heavy snow. Electronic waves passing between transmitters and receivers can detect obstacles in a similar manner to the optical beam method. However, optical beam method cannot detect pedestrians.
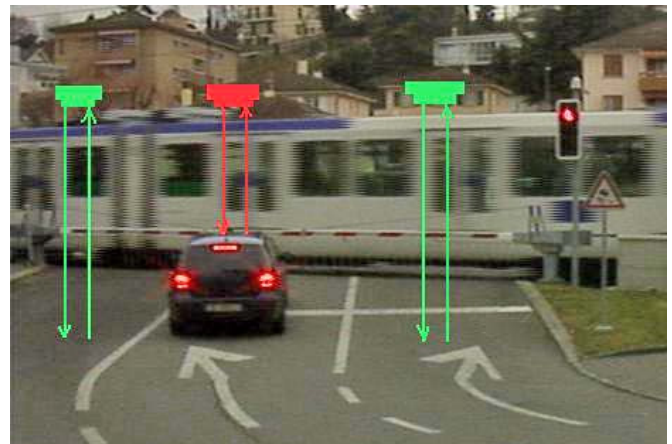

Fig. (2). Ultrasonic method.

Another technique consists in using ultrasonic detectors (Fig. **2**) which rely on differences in ultrasonic reflection times for detection. They transmit pulses of ultrasonic energy towards the roadway. The pulse is then reflected back more quickly when a vehicle passes through. As an advantage, sonic detectors can detect both stationary and moving vehicles. However, the disadvantages reside in their price and installation cost. Additionally, they are extremely sensitive to environmental conditions (inaccurate in congested conditions).

Another kind of obstacle detection system has recently been introduced on the railroad crossing such as radar device (Fig. **3**). The principle of this technique is as follows: Microwaves are sent from a transmitter based on the side of the roadway. The microwaves are reflected back to a receiving antenna with a different frequency. This change is picked up and reflects the presence of a moving object due to

the Doppler effect. Radar presents a lot of advantages such as the fact that traffic does not need to be disrupted for installation and it is immune to electromagnetic interference. The disadvantage is that it is hard to maintain.
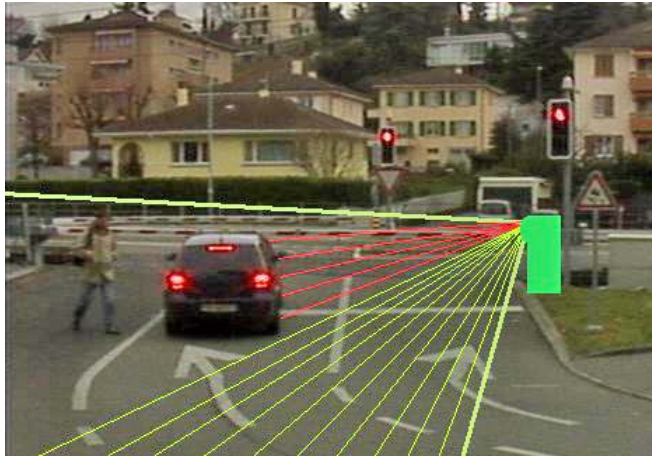

Fig. (3). Radar method.

The inductive loop is probably the most common form of vehicle detection. A wire is embedded under the roadway. A magnetic flux is generated around it. When a vehicle is passing over the wire, the flux is cut causing an increase in inductance, and the detection of an obstacle. This is what happens when a car pulls up to the intersection. The huge mass of metal that makes up a car alters the magnetic field around the loop, changing its inductance. This system is easy to install and is not subjected to environmental conditions. The main drawbacks are that it cannot detect pedestrians or bicycles, and the high cost of installation and maintenance, because it needs a large number of loops to be efficient.

Another object detection system has recently been designed by Hisamitsu [4] based on a 3D lidar (Light Detection and Ranging) technique in which the principle is the following: 3D lidar emits a laser pulse to an object, and measures the time that it takes for the reflected laser to return to the receiver (time-of-flight method). It acquires the distance from that object. Fig. (**4**) shows the object detection method used by the 3D lidar. A laser pulse is emitted in a way that it scans the entire area of a level crossing in the horizontal and vertical directions. The 3-D coordinate's values of each point measured from the reflected laser are returned to the 3-D lidar. Summarizing, optical beam and sonic detectors are definitely too expensive, not accurate enough and they need a high number of sensors.
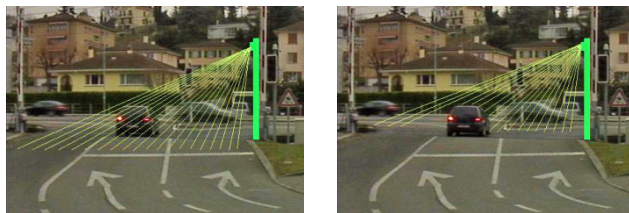

Fig. (4). Laser method.

They are also too sensitive to environmental conditions. The most relevant technologies for object detection at level crossings are radar, inductive loops and video imaging. The problem with inductive loops is that they detect only metal objects. Nowadays, many vehicles use carbon and aluminum

fiber, which are not well detected by the system. Afterwards, the maintenance of such a system is hard because it is located below the road. That is why we are focusing on video and now present some realizations and utilizations of these technologies.

3. VIDEO IMAGING ON LEVEL CROSSINGS

One of the main operational purposes for the introduction of CCTV (Closed Circuit Television) at LC is the automatic detection of specific events. Actually, no conventional vision-based system is installed at level crossings today, but a few level crossings are equipped by a camera-based system and are under test in Japan. The purpose is not to visually analyze such crossing zone when there is CCTV. An automatic process must be established in order to only transmit relevant information to the control room and train driver like the presence of an obstacle in the crossing zone at a critical time. A typical example of a hazardous scenario consists of detecting a car stationary in the crossing zone for a few minutes, this before the lowering of barriers and without a priori about the position of train. The decision that must be taken for a given situation depends on several factors such as its speed when the brakes are applied [5], the geography of the track, in particular the track slope the train travels over from when the brakes are applied to where the front of the train stops, the mass distribution of the train and the proximity of the nearest train to the level crossing. The automatic video system can have a priori about the position of the nearest train; this accurate information can be given permanently by the control room. Otherwise, another way is to rely on vision by detecting the movement of barriers. It is envisaged that the new technologies will increase the number of monitored camera images, give the staff earlier warnings to be able to produce more timely responses, and support the recording and collection of evidence of detected events. The level crossing obstacle detection system that uses cameras detects any kind of objects, such as pedestrians and wheelchairs, that cannot be detected by the conventional systems of the photoelectric (i.e. optical beam), ultrasonic, and loop coil types. These particular objects that can be detected using passive vision may be tracked close together in space and time. This allows us to model spatiotemporal behavior from image sequence. Nowadays, most of the conventional object detection systems described above supervises only dangerous crossing areas. According to the technology used, the detected object is tracked when it appears in the critical zone. For this purpose, using artificial vision can provide more relevant information and with such a system, dangerous scenarios can be recognized at the right time. Fig. (**5**) presents a reference scenario of hazardous situations detected from a stereo camera. Interpreting this scenario allows us to deduce the importance of the level crossings environment, which gives significant information such as road traffic and the status of the red-light. This additional information can be provided as input to the computational system related to the cameras.

3.1. Overview of Existing Systems

There were some recent works which use the artificial vision for controlling safety at railroad crossings [6], at railways [7] and for vehicle detection at road [8]. Referring to the literature, little research has focused on passive vision

to solve the problems at level crossings. Two mains systems based on CCD cameras can be distinguished:

- A level crossing obstacle detection system using one single camera: This system using video cameras has been developed by Foresti [9]. This system uses one single CCD camera placed on a high pole in a corner of the level crossing. First, the object detection is computed based on the difference between current and background images. Then, with the knowledge of intrinsic camera parameters, calibration matrix and ground plane hypothesis, the 3D position of the different objects found are computed. Then objects are tracked with an Extended Kalman Filter (EKF). Finally, object classification based on morphological statistical spectrum is performed in order to classify objects as car, bike, truck, pedestrian, dog, paper, etc. Tests proved that the system works well in different situations (different camera points of view, bad environmental conditions, noise, object occlusions). But this system proved to be limited in low illumination and the presence of shadows can lead to false alarm.

- A level crossing obstacle detection system using stereo cameras: M.Ohta [10] has developed a level crossing obstacle detection system using stereo cameras. The problem of using a single TV camera acquiring a wide-area image is car headlights or shadows, which cause errors in detection. The idea of Ohta is to use a stereo camera whose principle will be described later. The object is then projected on each TV screen with a disparity and, with some computation; a 3D shape of the object is obtained. It reduces false detections of car lights or shadows. Then the disparity image (which shows the distance of an object against the camera) is computed. Finally, the 3D shape of the whole scene is extracted and compared with the background shape. The obstacle detection using stereo cameras can detect both vehicles and pedestrians. The tests have proved that the system correctly detects objects during day and night under general weather conditions. But the main problem with this system (and with image processing in general) is that it is extremely sensitive to adverse weather conditions, like heavy rain, fog or snow.

3.2. Proposed System Design

In order to overcome the limits described above, the passive vision principle as shown in Fig. (**6**) will be adopted. For now, there are several manners for putting the cameras. This depends on the type of each level crossing and on the most potential side to be monitored. Currently, our proposed system consists of the use of two cameras, having a baseline of 40 centimeters, supervising the crossing zone. The cameras are placed so that the dangerous zones are well

Fig. (5). An example of hazardous situation scenario detected from cameras devices. (**a**) Pedestrians near a crossing, (**b**) appearance of a car, (**c**) A car approaching the crossing (first level of red-light sensor), (**d**) A car approaching the crossing (second level of red-light sensor), (**e**) A car on the crossing (third level of red-light sensor: a mandatory stop), violation of red-light, (**f**) A car stopped on the critical area, (**g**) Start of lowering barriers, (**h**) Car driver is still on the dangerous zone, (**i**) A barrier is lowered and jammed on the car, (**j**) Car driver has forced the barrier: It is full track, (**k**) The car passing the hazardous area, (**l**) A free and safe level crossing (barriers are still lowered).

(fully) monitored. The processing is based on the following two steps: Foreground extraction from image sequences, and 3D localization of detected moving foreground. For motion object segmentation, the background subtraction method is implemented. This technique proves effective for outdoor image sequences, especially with changes in brightness. The moving objects detected with a single camera will be taken into account for the 3D localization module. This relevant information will be useful for localization and recognition processes. To have accurate information for 3D localization, the stereo method is only applied to particular points of the moving object. The newly developed system was tested with a set of stereoscopic sensor in order to cover the whole area of a double-track level crossing. The rest of this paper will focus on the two previous techniques: foreground extraction and 3D localization.

Fig. (6). A level crossing (in Ecublens-Switzerland): obstacle detection system that utilizes stereo cameras and image processing technology.

4. FOREGROUND EXTRACTION

4.1. Moving Object Detection

In video surveillance, detection of moving objects from an image sequence is very important for target tracking, activity recognition, and behavior understanding. Motion detection aims at segmenting foreground regions corresponding to moving objects from the background. Background subtraction and temporal differencing are two popular approaches to segment moving objects in an image sequence under a stationary camera. Temporal differencing calculates the difference of pixel features between consecutive scene frames in an image sequence. It is very effective to accommodate environmental changes, but generally can only detect partial edge shapes of moving object. In this study, the first techniques family for foreground subtraction with background model updating strategies was adapted. Background subtraction is a very popular approach for foreground segmentation in a still scene. It detects moving objects in an image by evaluating the difference of pixel features of the current scene image against the reference background image. This approach is very sensitive to illumination changes unless adaptively updating the reference background. However, the background in a long image sequence is dynamic even if it is captured by a stationary camera. In this context, the

generalized mixture of Gaussians proposed in [11] has been used to model complex, non-static backgrounds. Methods employing a mixture of Gaussians have been widely incorporated into algorithms that utilize Bayesian frameworks [12], dense depth data [13], color and gradient information [14], mean-shift analysis [15], and region-based information [16]. A mixture of Gaussians does have some drawbacks though. Backgrounds having fast variations are not easily and accurately modeled with just a few Gaussians, which may fail to provide sensitive detection. In addition, depending on the learning rate to adapt to background changes, a mixture of Gaussians faces trade-off problems [17]. Recently, a new method has been proposed by Kim [18], who describes a technique for discriminating moving objects from the background. The adopted background subtraction scheme involves two stages, one for training and the other for detection.

The codebook algorithm adopts a quantization technique to construct a background model from long observation sequences. For each pixel, it builds a codebook consisting of different states of pixel based on color and intensity. Samples at each pixel are clustered based on a color distortion metric together with brightness bounds. The clusters do not necessarily correspond to single Gaussian or other parametric distributions. The background is encoded on a pixel-by-pixel basis. Detection involves testing the difference of the current image from the background model with respect to color and brightness differences. An incoming pixel must be classified into background or foreground class based on color and brightness distortion referring to background estimated model. Therefore, the chosen method for foreground extraction is suited to level crossing application. It is effective for motion detection under varying environmental conditions due to updating background. Moreover, the fundamental challenges is to achieve invariance to illumination changes, and more prominently, to shadows and highlights. The detected moving object includes the real object and its shadow. This over-segmentation will interfere the recognition and classification process which can cause therefore false alarms. It is to be noted that cast shadows (it's created by an object blocking the light source) is the type of shadow that we are interested in.

4.2. Shadow Removal

Dealing with shadows and highlights is essential in object detection and tracking applications such as automated video surveillance systems. Shadows occur frequently in a wide variety of scenes. It is more problematic for outdoor scenarios subject to variable lighting and weather conditions. However, this is undesirable due to the fact that they often lead to the result of irretrievable processing failures. For instance, the shadow cast by an object results in an improper segmentation result with serious artifacts, or detection of an imaginary object. This might result in shadows misclassified as objects or parts of objects due to the overestimation in a subsequent matching phase. Hence, the obtaining moving object is the sum of the real object and its shadows. We distinguish two types of shadows: Cast-shadows and self-shadows. Self-shadows have no influence on the size of object. This does not cause an over segmentation of moving regions. Only cast-shadows will be taken into consideration

for shadow removal. Cast-shadows refer to areas in the background projected by objects in the direction of light rays, producing distorted objects silhouettes. Two main techniques exist for shadow removal: model-based and property-based method. The first technique needs some a priori knowledge (geometry features of the scene, illumination direction, etc.). The second technique is the most important because it identifies shadows by exploiting their properties in brightness and color. Color information is exploited by means of several color spaces, and by means of photometric invariant features. These methods are mostly based on the fact that shadows change significantly the luminance but not the color. Several researches proposed shadow detection methods based on the RGB, HSV and YUV [19] color spaces. Most of research assumes that a shadow belongs to a dark region. By exploiting color information, we overcome the limitation of the dark regions classification process of [20], which assumes that cast shadow pixels are darker than self shadow pixels as an empirical criterion for shadow classification.

Thereby, another property of shadow can be explored since shadow does not alter the value of the invariant color like rgb or $c_1c_2c_3$. On the contrary, a material change modifies their value. This property has been explored by Salvador [21] using the $c_1c_2c_3$ photometric invariant color features. The identification of shadow pixels is achieved by analyzing the difference in the invariant feature values between the pixel to be classified and it correspondent in the background. Examples of photometric invariant features are Hue and Saturation in the HSV color space and the normalized-RGB color space (rgb) which is tested by Cavallaro [22]. We have adapted the technique proposed by Porikli [15] for shadow removal. Likelihood of being a shadow pixel is evaluated iteratively by observing the color space attributes and local spatial properties. The author carried out the evaluation of the proposed method in the RGB color space, and assumed that shadow decreases the luminance and changes saturation, yet it does not affect the hue. However, the detection of shadows is based on the fact that shadows change significantly the brightness of an area without greatly modifying the color information. The algorithm is designed to be able to work when the characterization of the camera, illumination and scene are unknown. Shadows cast on a surface reduce the surface intensities. It is important to be noted that no conventional methods for shadows removal which can be applied to any image sequences. Fig. (**7**) represents the results of moving object detection applying the codebook method described above that we have improved. Three image datasets for real environment at different level crossings are used for the experiments.

5. 3-D LOCALIZATION OF THE DETECTED OBJECT: OBJECT DEPTH ESTIMATION

Stereo vision has become one of the most extensively researched topics in computer vision. However, the use of one camera for object detection process as partly solved some problems related to object detection and recognition in various types of scenes. The use of multi-camera systems provides additional information, such as the depth of objects in a given scene. Dense or sparse stereovision techniques can be used to match points. In dense stereovision, all points of

an input image are taken into account for matching tasks. All pixels are used to prove the effectiveness of our proposed stereo matching algorithm compared with the state-of-art. This evaluation is done on the well known dataset [23] containing various pair of images with ground truth. The sparse stereovision is then adopted. Then, our proposed stereo matching algorithm is applied only on the moving zones detected automatically referring to the technique discussed in section 4. Each disparity, determined for each point of the scene, represents the coordinate gap of the same point between the two left- and right-hand images representing the same scene from two different points of view. A depth map is obtained from the two images. Each value of this obtained map is an estimation of the distance between a real point in the 3-D space and the stereoscopic sensor and is given according to its intrinsic and extrinsic parameters, such as the focal length and the baseline. When a point is imaged from two different viewpoints, its image projection undergoes a displacement from its position in the first image to that in the second image. The amount of displacement, alternatively called disparity, is inversely proportional to distance and may therefore be used to compute 3D geometry. Given a correspondence between imaged points from two known viewpoints, it is possible to compute depth by triangulation. The problem of establishing correspondence is a fundamental difficulty and is the subject of a large body of literature on stereo vision. One prominent approach is to correlate pixels of similar intensities in two images, using an assumption that each scene point reflects the same intensity of light in the two views. The work carried out by Okutomi and Kanade [24] extends this correlation approach to two, three or more images and demonstrates that using several cameras at different camera separations, or baselines, yields a significant improvement in reconstruction accuracy.

Most two-frame stereo matching approaches compute disparities and detect occlusions assuming that each pixel in the input image corresponds to a unique depth value. Until recently, computers were much too slow to even dream of having a real-time algorithm implementation of an algorithm based on color stereovision.

But costs and processing time are decreasing at a steady pace, and it is becoming realistic to believe that such a thing will be commonplace soon. The stereo algorithm presented later, springs from relaxation and energy minimization fields. Our approach aims to represent a novel framework to improve color dense stereo matching. As a first step, disparity map volume is initialized applying a new local correlation function. After that, an assessment of quality of matching is addressed. A confidence measure is attributed to all the pairs of matched pixels, which are classified into three classes. Then, the disparity value of all unclassified or badly-matched pixels is updated based only on stable pixels classified as well-matched. The selected pixels are used as input into disparity re-estimation modules to update the remaining points. Our main contribution is that we propose a novel approach to compute a confidence measure for each pair of pixels. The confidence measure is based on a set of original local parameters related to the correlation function used in the first step of our algorithm. This paper will focus on this point. The main goal of our study is to take into account both quality and speed.
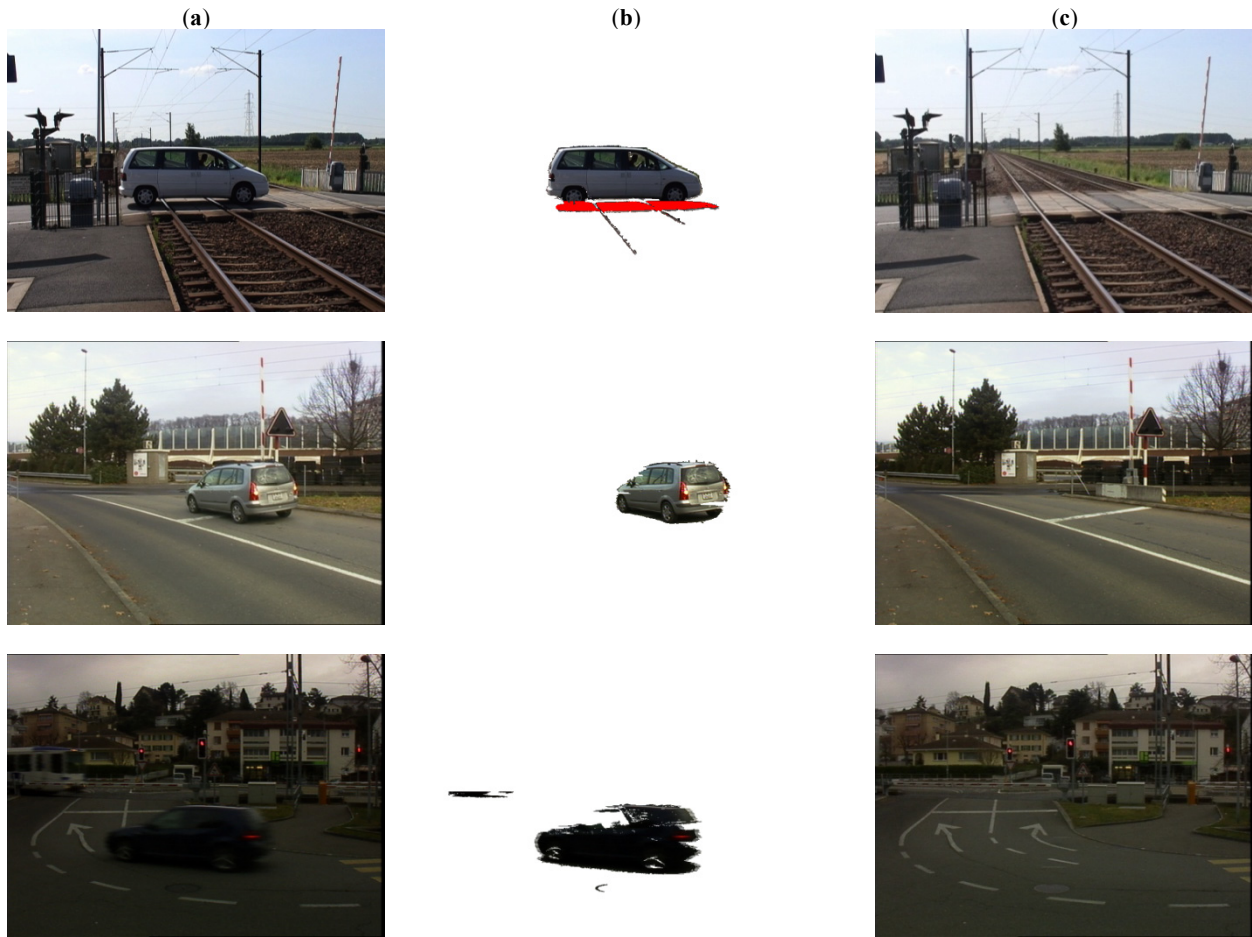
(a)   **(b)**   **(c)**

Fig. (7). Foreground extraction process. (**a**) Original image, (**b**) foreground extraction of moving object applying codebook method. The shadow is detected and represented by red color (**c**) the updated background.

Left image   Right image

Line i

$P_l(i,j)$ $P_r(i,v)$ $P_r(i,j)$

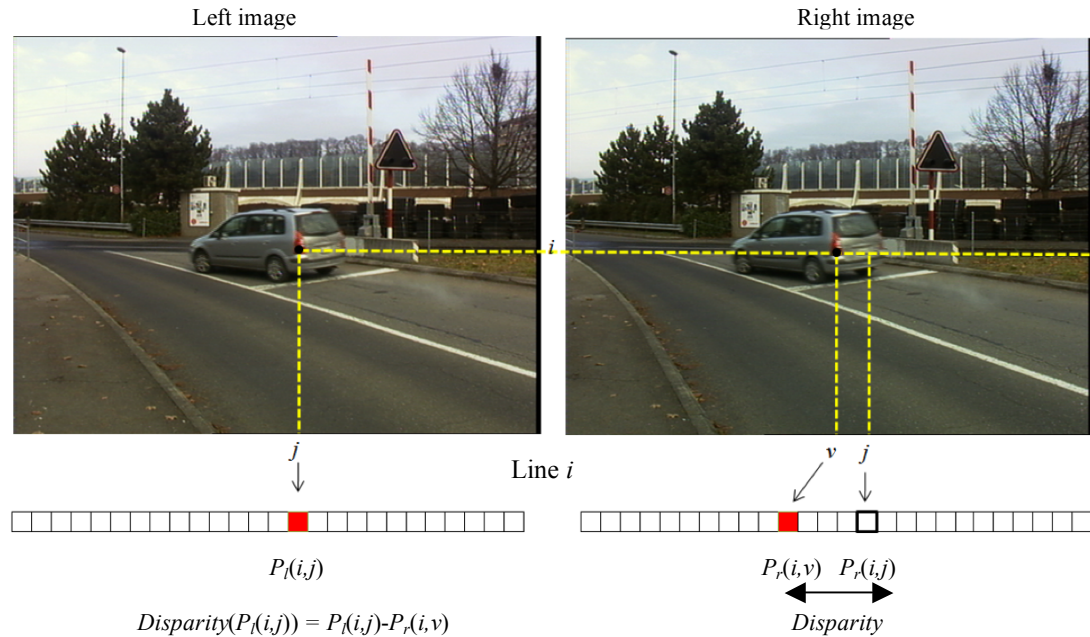$Disparity(P_l(i,j)) = P_l(i,j)-P_r(i,v)$ *Disparity*

Fig. (8). Stereo vision principle for estimation of depth foreground moved object. The horizontal yellow line means that the pixel to be matched and its corresponding in the other image have the same y coordinate. This configuration is obtaining after a mechanical calibration and a correction of optical aberrations. The vertical line of j coordinate in the left-hand image represents the position of the pixel to be matched. The same line is showed in the right-hand image in order to visually explain the concept of disparity. The vertical line of v coordinate in the right-hand image represents the corresponding pixel of the pixel to be matched. The gap between the j and v position is called "disparity".

5.1. Overview of the Proposed Stereo Matching Algorithm

We present, in this section, an overview of our stereo matching (the principle is shown in Fig. **8**) algorithm which allows us to improve the accuracy of disparity maps. The algorithm can be divided into three parts: Initialization of the disparity map, pixel classification, and disparity allocation. In the first step, we compute the correlation volume. Thus, we take advantage of local methods by applying a new color window correlation to build the correlation volume. It is called Weighted Average Color Difference (WACD) [25]. Firstly, this dissimilarity function is applied to all the pixels in stereo images using a non-adaptive square correlation window. This local method allows an initial appraisal of the disparity map. The second part of our algorithm classifies more accurately the matched pixels. Stereo matching accuracy may be affected by various factors including feature descriptors, similarity measures and matching approach. We assume that depth, generally derived from disparity, varies smoothly within color homogeneity in a given region. Depth discontinuities coincide generally with color boundaries or edges. Under the previous constraints, all matched pixels are classified into three categories according to their location: well-matched, badly-matched pixels and not-classified pixels. A new classification method based on a confidence measure approach is applied in this context. This confidence measure is computed for all matched pixels and is based on a set of local parameters referring to scores obtained from a new dissimilarity function. This method and the associated parameters are detailed later in this paper. This work is based on the principle of relaxation and the belief propagation method which are based on global criteria. However, the difference is that we consider only candidate pixels to evaluate the matched pair.

5.2. Weighted Average Color Difference Correlation Function

Disparity map is initialized applying a local method. The local approach consists in taking into account a set of neighborhoods pixels belonging to a support window centered on the pixels to be matched. Each pixel in the left-hand image has K candidate pixels in its corresponding right-hand image. Each pair of pixels, which is composed of both the pixel to be matched and a candidate pixel, is evaluated based on their neighborhoods pixels. We have exploited additional information such as color components to develop a new local correlation function. It is applied to each pairs of pixels and allows giving a dissimilarity measure. The candidate pixel, with which the correlation function gives the lowest correlation score, is kept as the best candidate. The lower the score, the better the matching. To deal with the image ambiguity problem, area-based local methods use, in most of cases, some kind of statistical correlation between color or intensity patterns in local support windows. This new correlation function is applied with a given support window. Neither the size nor the shape is treated in this study. Among the state-of-art local stereo matching function, we can mention the well-known algorithm named Sum of Absolute Differences (SAD) [26] which employs all neighborhoods pixels belonging to support windows. The particularity of our proposed correlation function is that we take into consideration only lines belonging to the used support window, these lines pass through the central pixel of the support window. Lines taken into account are the horizontal, the vertical and the two diagonals. The principle is explained in Fig. (**9**):
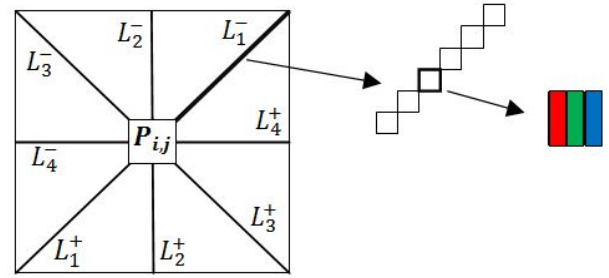

Fig. (9). Segments taken into account by the proposed correlation function to compute a dissimilarity measure between two support windows.

The lines are divided into segments and for each among them the average of each color component is computed.

Let denote L_n, $n = 1, 2, 3, 4$ (respectively L'_n) lines through the central pixel in the left-hand image (i.e. right-hand image) having (i, j) coordinates. We take into account as attributes all color components to compute the correlation score.

Let denote also $\Psi_{i,j} = \{ L^+_n, L^-_n \}$ (respectively $\Psi'_{i,j} = \{L^+_n, L^-_n\}$) a set of segments of each lines located on both sides of the central pixel belonging to left (respectively right) support window. In the first time, we are computing the average of each color component for each segment. These average colors are stored as a vector denoted as $v'_k = \bar{\psi}'_{i,j}$, $k = r, g, b$ (respectively $v'_k = \bar{\psi}'_{i,j}$).

Moreover, we have defined a measure denoted D which represents the distance difference between the k component of the central pixel and an average of the same component of a segment v_k. This measure is given by the equation 1:

$$D = MAX\left(\left|\left(P_{i,j,k} - v_k\right) - \left(P'_{i,j,k} - v'_k\right)\right|, 1\right) \qquad (1)$$

where $P_{i,j,k}$ is the k color component of the pixel of coordinate (i, j). The proposed distance D is weighted by a second measure denoted W which represent the difference between the two quantities δ_l and δ_r. (respectively δ_r) representing the average of each color component k for a given segment v_k of the left-hand image (respectively right-hand image). It is defined as follows:

$$W = MAX(|\delta_l - \delta_r|, 1) \qquad (2)$$

where $\delta_l = \dfrac{(P_{i,j,k} + v_k)}{2}$ and $\delta_r = \dfrac{(P'_{i,j,k} + v'_k)}{2}$

For all primary system such as RGB color space, it has often been proposed notably by Crouzil [27] to use the Euclidian distance to calculate color difference between two pixels. We have adopted in our context this metric for $n = 2$. The color difference between pixel $P_{i,j}$ and $P'_{i,j}$ is expressed as:

$$\Delta^n\left(P_{i,j}, P'_{i,j}\right) = \left(\sum_{k \in \{r,g,b\}} \left(P_{i,j,k} - P'_{i,j,k}\right)^n\right)^{\frac{1}{n}} \qquad (3)$$

This correlation function can be seen as weight average color differences (WACD) and it is given by the following equation:

$$DCMP\left(P_{i,j,k}, P'_{i,j,k}\right) = \qquad (4)$$

$$\Delta^2\left(P_{i,j,k}, P'_{i,j,k}\right) * \sum (D * W)$$

Fig. (**10**) illustrates disparity maps and error maps obtained for some images from the Middlebury dataset (Cones, Teddy, Venus, and Sawtooth).

5.3. Local Confidence Measure Estimation Theory

Related Works

In dense stereovision, several well-known stereo algorithms compute an initial disparity map from a pair of images under a known camera configuration. These algorithms are based loosely on local methods, such as window correlation, which take into account only the neighborhood of the pixel to be matched. The disparity map obtained has a lot of noise and erroneous values. This noise concerns mostly the pixels belonging to occluded or textureless image regions. An iterative process is then applied then to the initial disparity map in order to improve it. These methods use global primitives. Cost-relaxation approaches, which were invented by Marr and Poggio [28] and are picked up again by Brockers [29], belong to this family. Some research used a graph-based method [30] and color segmentation based stereo methods [31] which belong to so called "global approaches". Other approaches were proposed based on a probabilistic framework optimization, such as expectation-maximization [32] and belief propagation [33, 34]. These methods aim to obtain high-quality and accurate results, but are very expensive in terms of processing time. It is a real challenge to evaluate stereo methods in the case of noise, depth discontinuity, occlusions and non-textured image regions.

Besides, some mixture of local methods is first used to obtain an estimated disparity map which is improved in a second step using global methods. On the basis of local methods, a pixel on the left-hand image is evaluated with candidate pixels on the corresponding right-hand image. Yoon [35] applies window-based methods in the Lab color space coupled with an adaptive window [36] which tries to find an optimal support window for each pixel. These techniques assume that the neighborhood of a pixel to be matched presents homogeneity in terms of disparity values.
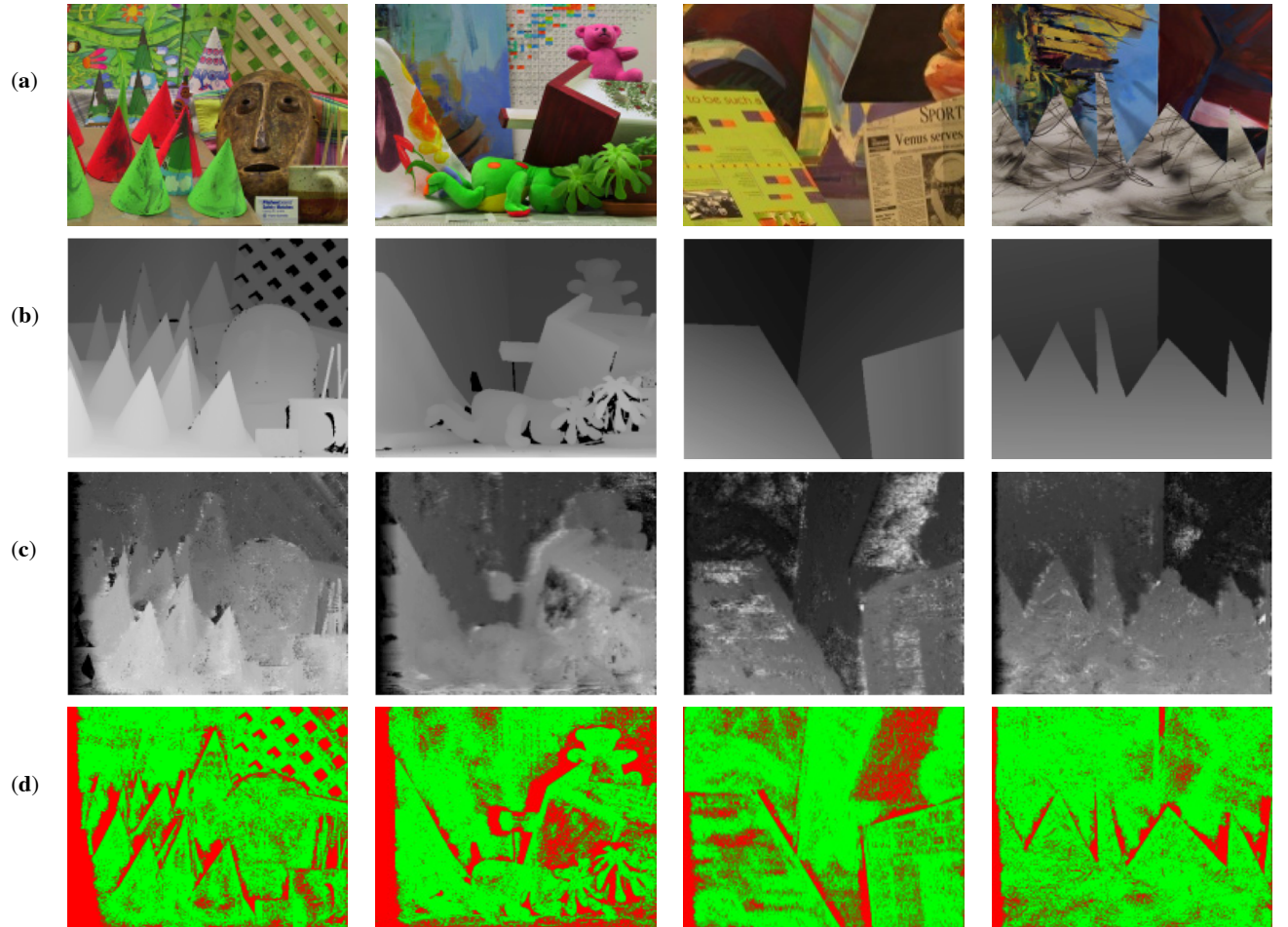

Fig. (10). Data sets and output images. (**a**) Left-hand images used for evaluation (**b**) ground truth disparity (**c**) Disparity maps obtained applying only dissimilarity function (**d**) Well- and badly-matched pixels: Green color for well-matched pixels and red color for badly-matched pixels.

In other words, all the pixels in the given correlation window must have very similar disparities.

Description of the Proposed Approach

In our approach, in order to reduce the processing time and deal with the problems of ambiguity in the matching process, the correlation function used to evaluate each stereo pair will only be applied on high color variation regions of the images.

The main idea of our approach is to compute a confidence measure for every matched pixel. Furthermore, confidence measure Ψ can be seen as a matching probability for pixel P_l (a pixel P in the left-hand image) with pixel P_r (a pixel P in the right-hand image), given some parameters. The way in which confidence measures are calculated is provided by Equation 5.

$$\psi\left(P_l^{i,j}, P_r^{i,v}\right) = P\left(P_l^{i,j}/P_r^{i,v}, N, min, \sigma, \omega\right) \qquad (5)$$

The confidence measure with its parameters is given by Equation 6:

$$\psi\left(P_l^{i,j}, P_r^{i,v}\right) = \left(1 - \frac{min}{\omega}\right)^{N^2.log(\sigma)} \qquad (6)$$

where:

<u>min: The Best Correlation Score</u>

The output of the dissimilarity function is a measure representing the degree of similarity between two pixels. Then, the candidate pixels are ranked in increasing order according to their corresponding scores. The couple of pixels that has the minimum score is considered as the best-matched pixels. The lower the score, the better the matching. The nearer the minimum score to zero, the greater the chance of the candidate pixel being the actual correspondent.

<u>N: Number of Potential Candidate Pixels</u>

This parameter represents the number of potential candidate pixels having similar scores. N has a big influence because it reflects the behavior of the dissimilarity function. A high value of N means that the first candidate pixel is located in a uniform color region of the frame.

The lower the value of N, the fewer the potential candidate pixels. In the case where there are a few candidates, the chosen candidate pixel has a greater chance of being the actual correspondent. Indeed, the pixel to be matched belongs to a region with high variation of color components.

While establishing the relationship between N and min values, with a very small value of N and a minimum score min, near to zero for instance, the pixel to be matched probably belongs to a region of high color variation.

<u>σ: Disparity Variation of N Pixels</u>

A disparity value is obtained for each candidate pixel. For the N potential candidate pixels, we compute the standard deviation σ on the N disparity values. A small σ means that the N considered pixels are neighbors. In this case, the true candidate pixel should belong to a particular region of the frame, such as, an edge or a transition point.

Therefore, it increases the confidence measure. A large σ means that the N candidate pixels taken into account are situated in a uniform color region.

<u>ω: Gap Value</u>

This parameter represents the difference between the Nth and $(N+1)^{th}$ scores given with the dissimilarity function used. It is introduced to adjust the impact of the min score.

To ensure that this function gives a value between 0 and 1, some constraints are introduced. The min parameter must not be higher than the ω one. If so, parameter ω is forced to $(min + 1)$. However, the $log(\sigma)$ term is used instead of σ alone. It has a big influence on the confidence measure in case of high values of σ, and it is indifferent otherwise. This leads to reducing the impact of high values of σ and to obtaining coherent confidence measures.

The number N of potential candidate pixels is deduced from the M scores obtained with the dissimilarity function described in section 5.2. The main idea is to detect major differences between successive scores. These differences are called main gaps. Let f denote a function which represents all scores given by the dissimilarity function in increasing order. Then, we apply the average rate growth to the f function. This second function can be denoted by η and can be seen as the ratio of the difference between a given score and the first score, and the difference between their ranks. This function is defined in Equation 7.

$$\eta(x_m) = \frac{f_{x_m}^{i,j} - f_{x_1}^{i,j}}{x_m - x_1} \; m = 1 \dots k \qquad (7)$$

where $f_{x_m}^{i,j}$ is the mth score among the M scores of the pixel of (i, j) coordinate. x_m is the rank of the mth score.

$$\xi(x_m) = \frac{\eta_{x_m}^{i,j} - \eta_{x_{m-1}}^{i,j}}{m^2} \; m = 1 \dots k \qquad (8)$$

The previous function (Equation 8) is used to characterize jump scores and is applied only in the case where $(\eta_{x_m}^{i,j} - \eta_{x_{m-1}}^{i,j})$ is a positive value. We have introduced parameter m^2 in order to penalize candidate pixels according to their rank. The number of potential candidate pixels is given by Equation 9.

Fig. (**11**) illustrates an example of the function (8) computed for a well-matched pixel.

$$N = \underset{m}{\text{Argmax}} \; \xi(x_m) \qquad (9)$$

6. EXPERIMENTAL RESULTS

In this section, we'll discuss and evaluate the performances of our proposed approach. On the first hand, this evaluation is done on the well-known conventional stereo images with available ground truths [23] to test the relevance of the accuracy of our approach. In this evaluation, four pairs of stereo images with ground truth are used: Cones, Teddy, Venus and Sawtooth. These pairs of images are chosen so that they are very different from each either. One contains many occluded regions; other contains regions

of uniform color. Also, these images are the most used for evaluation of stereo matching algorithms. As a first step, the disparity map is initialized by applying the dissimilarity function proposed in [25]. This provides a first visual rendering of the disparity map for Cones (Fig. **12**).

The used dissimilarity function has the particularity of matching well most pixels in regions having high variation of color components. As shown in Fig. (**12d**), Pixels belonging to uniform color regions or depth discontinuity regions are badly-matched. This can be explained by the presence of several potential candidate pixels of the pixel to be matched. Therefore, this increases the error of the matching task. Our stereo matching algorithm is first applied to the whole pixels in the image. This allows us to obtain a dense disparity map.
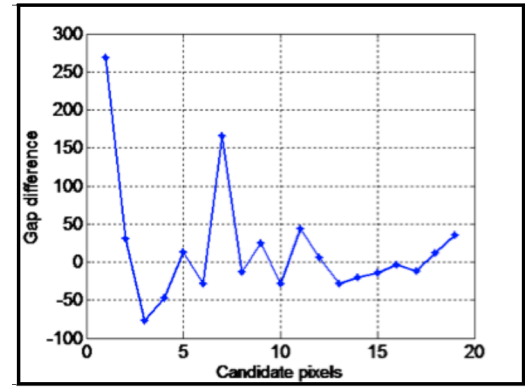

Fig. (11). The number N of potential candidate pixels is the rank of global maximum of ξ function.
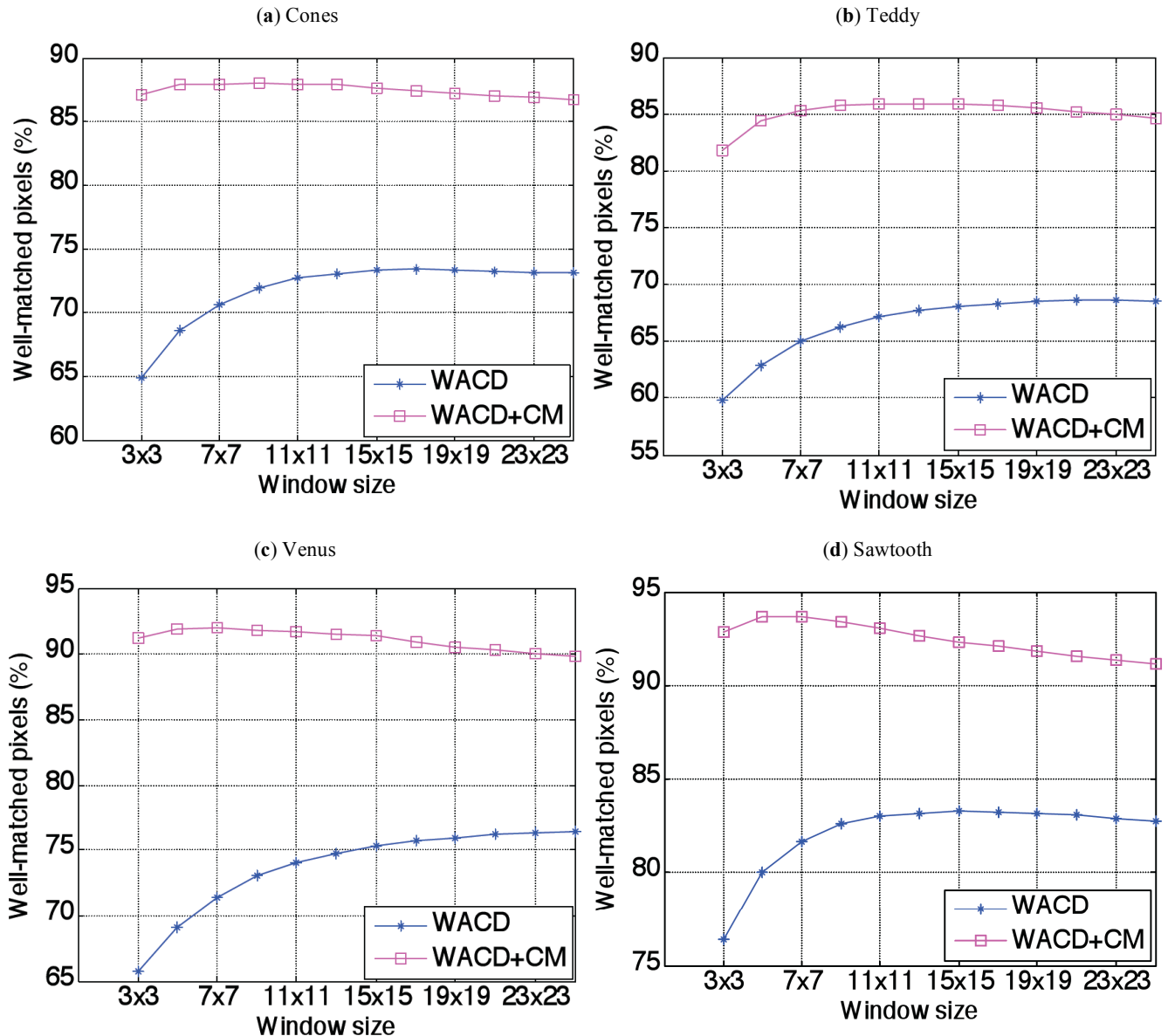

Fig. (12). Well-matched pixels rate using only WACD correlation function (blue curves) described in section 5.2, and introducing both confidence measure approach and homogeneous color regions elimination (pink color) described in section 5.3.
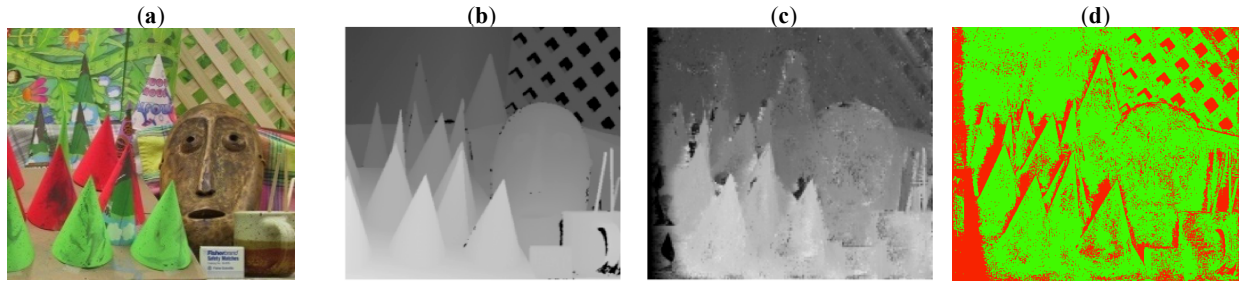
Fig. (13). Left-hand "Cones" image and output images. (**a**) Left-hand images used for evaluation (Cones) (**b**) ground truth disparity (**c**) Disparity map obtained with WACD dissimilarity function application only (**d**) Well- and badly-matched pixels: Green color for well-matched pixels and red color for badly-matched pixels.
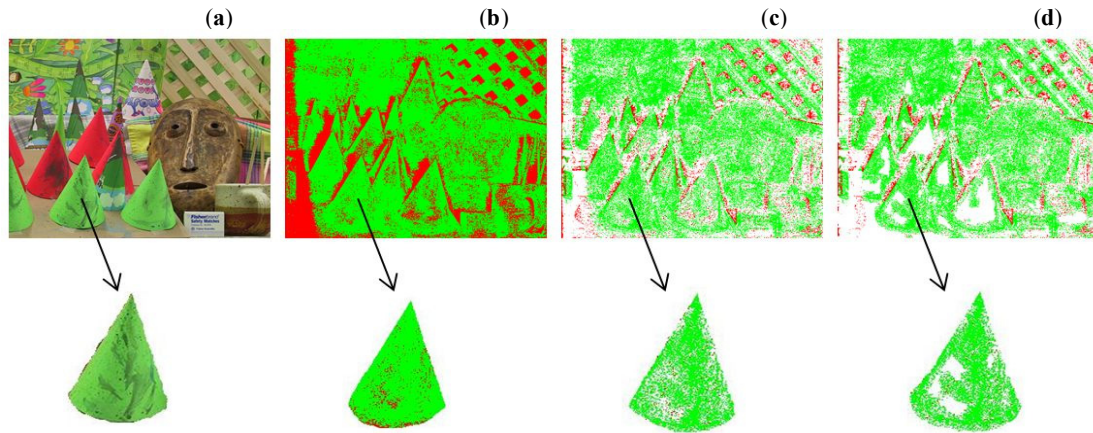

Fig. (14). (**a**) Left-hand Cones image (**b**) Well-matched and badly-matched pixels with WACD dissimilarity function application only (**c**) Well-matched in green, badly-matched in red and unclassified pixels in white for a confidence measure of 96% (**d**) After elimination of pixels belonging to regions having a low color variation.

As a second step, and in order to quantify and to automatically identify well-matched pixels from badly-matched pixels, we have exploited the confidence measure method described in section 4. Each couple of matched pixels is evaluated and belongs to one of the following three categories for a given confidence measure threshold T:

- Well-matched pixels having a confidence measure higher than T,

- Badly-matched pixels having a confidence measure higher than T and an erroneous disparity value according to the ground truth image,

- Unclassified pixels having a confidence measure lower than T.

The disparity map is obtained with an optimal window size (11x11) defined empirically. All matched pixels are classified into the three previous categories. Global results are shown in Fig. (**12**). According to Fig. (**13d**), the badly-matched pixels represent either occluded pixels or pixels belonging to uniform color regions. In order to reduce the impact of error matching and to have a high gap between the rate of well-matched and badly-matched pixels, only pixels belonging to regions of high color variation are considered. A pixel will be matched only if the sum of all color component variations of neighborhoods pixels belonging to a support window is higher than a threshold. This threshold is computed based on dynamics of color in the image.

In Fig. (**12**), four diagrams are used to illustrate the matching performances using the WACD correlation function for the lower curves, and using the same function with introduction of confidence measure for the upper curves. We can notice that the use of confidence measure applied to matched pixels allows us to improve the matching rates significantly. Thus, for a correlation window of 11x11 pixels and a 96% confidence measure, the good matching rate passes from:

- 73% to 88% for cones,

- 67% to 86% for Teddy,

- 74% to 92% for Venus,

- 83% to 93% for Sawtooth.

The rate of well-matched pixels is initially evaluated by comparison of the estimated disparity and the corresponding disparity values of ground truth images. This concern all matched pixels of the image. The initial rate of well-matched pixels is compared with well-known stereo matching algorithms such as Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD). Our algorithm show an improvement of 10% of well-matched pixels rate with a 5x5 window size, and for about 6% with 11x11 window size.

In Fig. (**14b**), for the extracted cone, we can notice that the number of badly-matched pixels (in red) is quite high. These pixels are located in uniform regions of the cone. The

application of the confidence measure (passage from **14b** to **14c**) leads to an important decrease in the number of badly-matched pixels. In fact, in Fig. (**14c**) we can notice that the number of red pixels has reduced. The white pixels in Fig. (**14c**) are the ones with a confidence measure lower than the given threshold (96%). In Fig. (**14d**) the unclassified pixels belonging to homogeneous regions in terms of color are identified and marked in white. Finally, in our method, we have only considered pixels having a high confidence measure.

The introduction of the confidence measure principle leads to decrease the rate of badly-matched pixel that will be considered as unclassified pixels. Moreover, this increases consistently the rate of well-matched pixels for a given region. This rate is calculated based only on well- and badly-matched pixels. A significantly rate of well-matched pixels in a given region (half of pixels minimum) not allows the re-estimation of the remaining pixels which are labeled as unclassified pixels.

It is to be noted that the rate of well-matched pixels, obtained with the use of the confidence measure, concerns only pixels belonging to high color variation zones (around 80% on average for the four pairs of images tested). The remaining pixels, those belonging to uniform zones in terms of color, could be assessed if need. In Fig. (**14**), the different steps of our approach are illustrated. We have extracted manually a cone to illustrate visually the improvement according to each step. For real applications, moving zones are extracted automatically according to the motion detection technique described in the section 4. For a given moving region, we assume that the disparity values of each pixel are homogeneous and very similar. For this raison that the badly-matched pixels will be updated only if the rate of badly-matched pixels is less than 20% of pixels in a given zones. The sufficient number of well-matched pixels leads to a better estimate of the 3-D position of a given object.

We have experiment our algorithms in real images acquired from two level crossings in Switzerland. These datasets contains a lot of real scenario with cars, pedestrians and objects stationary. Background changes detection step is first performed on the left-hand images by applying background subtraction technique described in section 4. Therefore, only on the obtained moving zones that the stereo matching algorithm is applied. An initial sparse disparity map is obtained. Then, a confidence measure is attributed to each pair of matched pixels. Through this step the badly-matched pixel are automatically highlighted. Fig. (**15**) shows the result of the use of the three steps described above: background subtraction, depth initialization and matching error elimination.

The introduction of the confidence measure technique results in a very accurate disparity value for each segmented object. Referring to the disparity values of each point in each segmented object it is now possible to best estimate, according to the triangulation procedure, the 3-D position and spatial occupancy rate in the scene of each segmented object. The transformation of an image plane point $p = (x, y)$ into a 3-D reference system point $P = (X, Y, Z)$ must be performed. The real distance of a real point is calculated by the triangulation technique that depends on a number of intrinsic and extrinsic parameters, such as the distance f between lens and film called camera focal lengths and the offset distance b between cameras. These two parameters are the same for both images in a given configuration. Hence, the distance D of a point in the real word depends essentially on the disparity parameter and is given by the Equation 10:

$$D = bf/d \qquad (10)$$

where D represents the 3-D distance between the camera and a real point and d is the disparity value. An accurate 3-D localization is very useful to best localize each object in order to understand their activity around the level crossing.
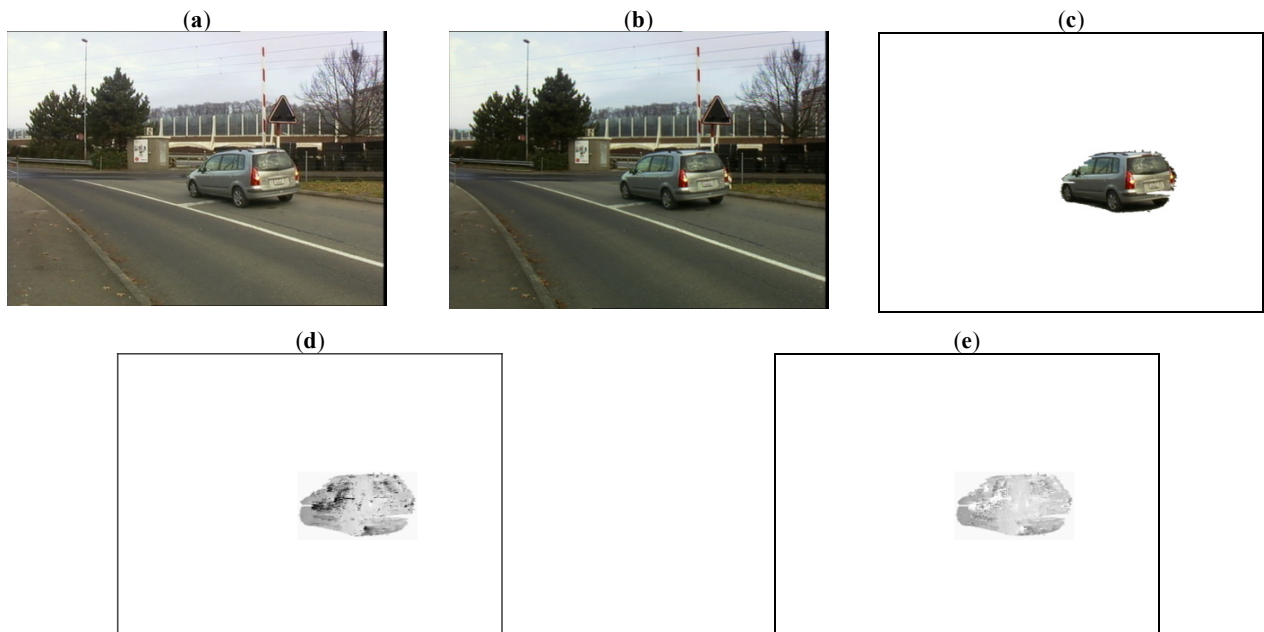

Fig. (15). An example of the disparity map obtained for each moving object (**a**) Left-hand image (**b**) Right-hand image, (**c**) The extracted moving object, (**d**) Disparity map obtained applying the proposed (WACD) matching function, (**e**) Improved disparity mapby introducing confidence measure technique.
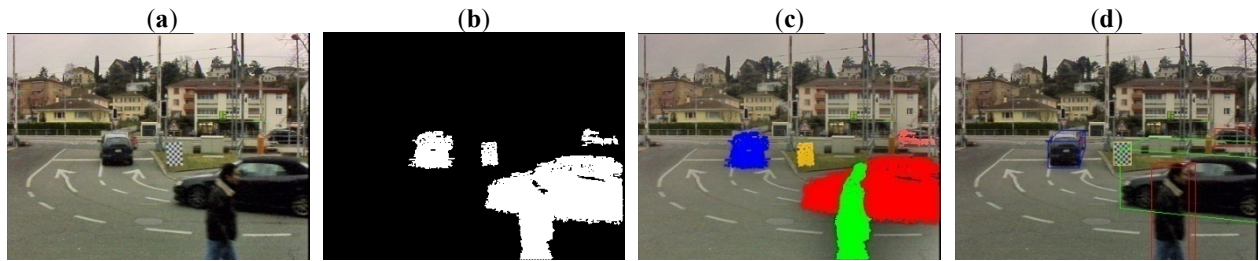
Fig. (16). This example shows the segmentation result (**a**) original image (**b**) a binary image showing the moving objects (**c**) Object localization obtained according to the stereo matching principle, (**d**) 3-D object localization.

Although the processing time increases more by applying the stereo matching process, the localization results are more accurate.

Fig. (**16**) highlight the need to use an additional video sensor. It is an image captured from a scenario containing a set of moving objects such as pedestrian and cars and an object stationary. Notice that a pedestrian and a car are considered as the same moving object. The stereo matching process allows us to separate the overlapping targets which are previously segmented. This separation is performed according to the 3-D distance of each point. This gives accurate features which facilitate to effectively recognize and afterward track each target.

On the other hand, in the foreground extraction process, motion could be used to re-estimate the remaining non-matched pixels. For a given moving region, it is reasonable to consider that the disparity values will stay homogeneous. Then, the 3-D localization is carried out on only moving regions. A way to re-estimate these remaining disparities is to consider for instance, the modal class disparity of well-matched pixels.

7. CONCLUSION AND PERSPECTIVE

A background subtraction technique for detection of moving objects was applied on natural images sequences. The used non-parametric technique entitled Codebook proves its effectiveness in case of small illumination variation in the scene but fails with a fast illumination background changes. To overcome this problem and to meet the requirements related to level crossings safety, we are developing a new approach that is based on Independent Component Analyses (ICA) technique. The latter is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It aims to find the source signals from observations data. The observations matrix can be formed by both a random background image and other containing arbitrary foreground objects. The estimated sources signals correspond to the modeled background signal and foreground, which distinctly preserves the foreground object without the detailed contents of the reference background [38].

A novel method was also proposed in section 5 of this paper for 3-D object localization. This aims to obtain an accurate disparity map for 3D localization of objects in level crossing. We have applied a local stereo matching method to initialize the disparity value for each pixel in moving zones. We have then introduced new local parameter in order to compute a confidence measure for each matched pixel. The

main contribution of our approach is twofold. On the one hand, new parameters introduced above can obtain important information for occluded and uniform region detection. On the other hand, unclassified pixels disparities can be updated in a post-processing step in order to obtain a more accurate disparity map. This depends on the disparity measure of each well-matched pixel classified using the confidence measure approach. The choice of the confidence threshold is very important for the success of 3-D estimation process of each moving object. This choice depends on the rate of well-matched pixels among all matched-pixels in a given region. Such a confidence threshold is considered as optimal when it satisfies the following function $T = argmax_T\ W_i$ in which W_i is the rate of well-matched pixels in the region i. In order to update the values of disparity of the pixels whose confidence measure is lower than a given threshold (those belonging to uniform regions), the values of disparity in the surrounding pixels can be used: segmentation, disparity modal class, clustering, etc. [37]. Another way is to consider the motion information to re-estimate unclassified pixels.

Our approach is highly dependent on the dissimilarity function used for computing the score of all matched pixels. However, a more extensive study will be carried out in order to enhance the used dissimilarity function. The rate of well-matched pixels will, therefore, be improved. The results are encouraging in terms of processing time which is compatible for a real-time implementation. The promising result obtained allows us to follow this track.

An extensive evaluation of our algorithms will be carried out considering real word situations. Datasets obtained from a series of level crossings in Lausanne (Switzerland) are used. Many real scenarios of cars, pedestrians, objects, crossing several different LC are included in the datasets. This will allow us to deeply evaluate the accuracy of our obstacle detection system in terms of objects extraction and then 3D localization. The next work consists to perform the tracking module. A vector of parameters will be constructed referring to 2-D and 3-D localization modules. The behavior of each target and their interaction will define the dangerousness of each situation. Of course, our detection system will be coupled with a communication one. These data can be sent to the drivers of the trains approaching the crossing, to inform them well in advance, on the existence of an obstacle in the dangerous zone and the status at the level crossings. In this case, the purpose also is to inform the car drivers in order to make the best of, the waiting period, in function of the situation. Furthermore, the condition of the level crossing, as well as the approach of the train can be transmitted to a control centre, either continuously, or at

regular intervals. Depending on the conditions at the crossing, the train driver can be warned to stop or to reduce the speed. This fits under study in the framework of the French national work program entitled PANsafer: towards a safer level crossing.

ACKNOWLEDGEMENTS

Our thanks go to TL ("Transports publics de la régionLausannoise"), which is the transportation company of the Lausanne city in Switzerland. By allowing the acquisition of images at several level crossings on their train lines, the TL allowed us progressing in our study, in particular for the evaluation of the proposed solutions."

REFERENCES

[1]   ERA (European Railway Agency), "A Summary of 2004-2005 EU Statistics on Railway Safety", Source of data: Eurostat.

[2]   Project SELCAT (Safer European Level Crossing Appraisal and Technology), A Co-ordination Action of the European Commission's 6th Framework Programme, 2006-2008.

[3]   Government's IT Strategy Headquarters: "New IT Reform Strategy", The realization of society, in which everyone can benefit from IT, anytime and anywhere. (In Japanese), January 19, 2006.

[4]   Y. Hisamitsu, K. Sekimoto, K. Nagata, "3-D Laser Radar level crossing obstacle detection system", IHI Engineering Review, vol. 41, no. 2, pp. 51-57, 2008.

[5]   D. Barney, D. Haley, and G. Nikandros, "Calculating train braking distance", Conference in Research and Practice in Information Technology Series, vol. 146, pp. 23-29, 2001.

[6]   I. Yoda, K. Sakaue, and D. Hosotani,"Multi-point stereo camera system for controlling safety at railroad crossings", IEEE ICVS, pp. 4-7, 2006.

[7]   J.J.D. Garcia, J.U. Urena, A.A. Hernandez, M.Q. Mazo, J.F. Vazquez, and M.-J Diaz, "Multi-sensory system for obstacle detection on railways", IEEE Instrumentation and Measurement Technology Conference Proceedings (IMTC), Victoria, BC, 2008, pp. 2091-2096.

[8]   G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and A. Broggi, "Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis", IEEE Trans Image Processing, vol.15, no. : 8, pp. 2364-2375, 2006.

[9]   G.L. Foresti, "A Real-time system for video surveillance of unattended outdoor environments", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 6, pp. 697- 704, 1998.

[10]   M. Ohta, "Level crossings obstacle detection system using stereo cameras", Quarterly Report of RTRI, vol. 46, no. 2, pp. 110-117, 2005.

[11]   C. Stauffer, and W. Grimson, "Adaptive background mixture models for real-time tracking", IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2246- 2252, 1999

[12]   D.S. Lee, J.J. Hull, and B. Erol, "A bayesian framework for gaussian mixture background modeling", IEEE International Conference on Image Processing, vol. 2, part III, pp. 973-976, 2003.

[13]   M. Harville, "A Framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models", European Conference on Computer Vision, part III, pp. 543-60, 2002.

[14]   O. Javed, K. Shafique, and M. Shah, "A Hierarchical Approach to Robust Background Subtraction Using Color and Gradient Information", IEEE Workshop on Motion and Video Computing (MOTION'02), pp. 22-27, 2002.

[15]   F. Porikli, O. Tuzel, "Human body tracking by adaptive background models and mean-shift analysis", IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS- ICVS, pp. 1-9, 2003.

[16]   M. Cristani, M. Bicego and V. Murino. "Integrated region- and pixel based approach to background modeling", proceedings of IEEE Workshop on Motion and Video Computing, pp. 3-8, 2002.

[17]   K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance", International Conference on Computer Vision, pp. 255-61. 1999.

[18]   K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Realtime foreground-background segmentation using codebook model", Journal of Real-Time Imaging, Special Issue on Video Object Processing, vol. 11, no. 3, pp. 172-185, 2005.

[19]   O. Schreer, I. Feldmann, U. Goelz, and P. Kauff, "Fast and robust shadow detection in videoconference application", in Proc. Of VIPromCom 2002, 4th EURASIP IEEE International Symposium on Video Processing and Multimedia Communications, Zadar, Croatia, 2002, pp. 371-375.

[20]   C. Jian, and M. O. Ward, "Shadow segmentation and classification in a constrained environment", CVGIP: Image Understanding, vol. 59, no. 2, pp. 231-225, 1994.

[21]   E. Salvador, A. Cavallaro and T. Ebrahimi, "Spatio-temporal shadow segmentation and tracking", Proceedings of SPIE's Image and Video Communications and Processing, vol. 5022, pp. 389- 400, 2003.

[22]   A. Cavallaro, E. Salvador, and T. Ebrahimi, " Detecting shadows in image sequences", in Proc. First European Conference on Visual Media Production, London (UK), 2004, pp.15-16.

[23]   D. Scharstein, and R. Szeliski, "Middlebury stereo vision research page, http://vision.middlebury.edu/stereo/

[24]   M. Okutomiand, T. Kanade, "A multiple-baseline stereo", IEEE transactions on pattern analysis and machine intelligence, vol. 15, no. 4, pp.353-363, 1993.

[25]   N. Fakhfakh, L. Khoudour, and M. El-Koursi, "Mise en correspondance stéréoscopique d'images couleur pour la détection d'objets obstruant la voie aux passages à niveau", in TELECOM'09 & 6ème JFMMA, Agadir, Maroc, 2009, pp. 206-209.

[26]   J. Martin, and J-L. Krowley, "Experimental comparison of correlation techniques", International Conference on Intelligent Autonomous Systems (IAS4), 2002.

[27]   A. Crouzil, "Perception du relief et du mouvement par analyse d'une séquence stéréoscopique d'Images.",University of Paul Sabatier, UPS, Toulouse, 1997, pp.58-60.

[28]   D. Marr, and T. Poggio,"Cooperative computation of stereo disparity", in American Association for the Advancement of Science, vol. 194, no. 4262, pp. 283-287, 1976.

[29]   R. Brockers, M. Hund, and B. Mertsching,"Stereo vision using cost-relaxation with 3D support regions", In Image and Vision Computing, New Zealand, 2005.

[30]   P. Foggia, J.M. Jolion, A. Limongiello, and M. Vento, "Stereo vision for obstacle detection: a graph-based approach", Lecture Notes in Computer Science, Springer Berlin/Heidelberg, 2007, pp. 37-48.

[31]   Y. Taguchi, B. Wilburn, and C. L. Zitnick," Stereo reconstruction with mixed pixels using adaptive over-segmentation", In CVPR, pp. 1-8, Anchorage, Alaska, 2008.

[32]   W. Xiong, H.S. Chung, J. Jia, "Fractional stereo matching using expectation-maximization.", IEEE TPAMI, vol. 31, no. 3, pp. 428-443, 2008.

[33]   C. Lee, and Y. S. Ho, "Disparity estimation using belief propagation for view interpolation", computers and comunications(ITC-CSCC), Japan, 2008, pp. 21-24.

[34]   J. Sun, N. -N. Zheng, and H.Y. Shum," Stereo matching using belief propagation", IEEE TPAMI, vol. 25, no. 7, pp. 787-800, 2003.

[35]   K.J. Yoon, S. Kweon, "Adaptative support-weight approach for correspondence search", IEEE TPAMI, vol. 28, no. 4, pp. 650-656, 2006.

[36]   O. Veksler, "Fast variable window for stereo correspondence using integral image", Proc. IEEE Conf. Computer Vision and Pattern Recognition, Madison, Wisconsin, 2003, vol. 1, pp. 556-561.

[37]   A. Klaus, M. Sormann, K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", ICPR, pp. 15-18, 2006.

[38]   D-M. Tsai, and S-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance", IEEE Transactions on Image Processing, vol. 18, no. 1, pp. 158-167, 2009.