Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2011, Article ID 548604, 15 pages
doi:10.1155/2011/548604

Research Article

3D Objects Localization Using Fuzzy Approach and Hierarchical Belief Propagation: Application at Level Crossings

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

[1] *Universitè Lille Nord de France, 59000 Lille, INRETS, Evaluation and Safety of Automated Transport Systems (ESTAS) and Electronic, Waves and Signal Processing Research Laboratory for Transport (LEOST), 59650 Villeneuve d'Ascq, France*
[2] *Laboratory of Microengineering for Manufacturing (LPM), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*

Correspondence should be addressed to N. Fakhfakh, nizar.fakhfakh@inrets.fr

Received 15 April 2010; Revised 7 September 2010; Accepted 1 October 2010

Academic Editor: Dan Schonfeld

Technological solutions for obstacle-detection systems have been proposed to prevent accidents in safety-transport applications. In order to avoid the limits of these proposed technologies, an obstacle-detection system utilizing stereo cameras is proposed to detect and localize multiple objects at level crossings. Background subtraction is first performed using the color independent component analysis technique, which has proved its performance against other well-known object-detection methods. The main contribution is the development of a robust stereo-matching algorithm which reliably localizes in 3D each segmented object. A standard stereo dataset and real-world images are used to test and evaluate the performances of the proposed algorithm to prove the efficiency and the robustness of the proposed video-surveillance system.

1. Introduction

In recent years, public security has been facing an increasing demand from the general public as well as from governments. An important part of the efforts to prevent the threats to security is the ever-increasing use of video-surveillance cameras throughout the network in order to monitor and detect incidents without delay. Existing surveillance systems rely on human observation of video streams for high-level classification and recognition. The typically large number of cameras makes this solution inefficient and in many cases unfeasible. Although the basic imaging technologies for simple surveillance are available today, the reliable deployment of them in a large network is still ongoing research.

In the context of railway transport, one of the major issues is the monitoring of linear infrastructures such as railway tracks and level crossings (LCs) which represent an interaction between a road and a railway track. The latter represents what is called extended perimeters. Numerous acts of malevolence occur in those areas. One can refer in particular to: objects hanging from catenaries, objects that may explode under the ballast, and obstacles at LCs. Transportation network could be interrupted aiming at causing economical damage or damaging a symbolic landmark.

The advanced surveillance system we intend to present here after relates to problems of safety and security at LCs. For some years, road and railways operators have shown growing interest in improving the safety and security of level crossings (LCs). They have been identified as a particular weak point in the safety of road and railway infrastructures. Road and highway safety professionals from several countries have dealt with the same subject: providing safer LCs. Recently, the EU's FP6 SELCAT project [1] (Safer European Level Crossing Appraisal and Technology) has provided recommendations for actions and evaluation of technological solutions to improve the safety at LCs. A new French project entitled PanSafer [2] aims at proposing such technologies, building upon the results provided by SELCAT. The present work is developed as part of PanSafer.

High-technology systems are developed so as to avoid collisions between trains and road vehicles. Nevertheless, high safety requirements may mean a costly systems which will hinder their actual use. Systems which have unacceptable

levels of false/missed detection have adverse effects and should not be implemented either. Some conventional object-detection systems have been tested at level crossings, and provide more or less significant information. Referring to the literature, little research has focused on passive vision to solve the problems at LCs. Among the existing systems, two of them based on CCTV cameras are to be distinguished: one of them is a system using a single camera [3]. It uses a single CCD camera placed on a high pole in a corner of the LC, classifying objects such as cars, bikes, trucks, pedestrians, dogs, and papers and localizing them according to the camera calibration, assuming a planar model of the road and railroad. This system is prone to false and missed alarms caused by fast illumination changes or shadows. The other one is a system using stereo cameras [4], with a stereo-matching algorithm and 3D background removal. This system more or less detects vehicles and pedestrians by day and night under usual weather conditions, but it is extremely sensitive to adverse weather conditions, like heavy rain, fog, or snow.

This paper is organized as follows: after an introduction covering the problem, we describe the requirements of the LC's application in Section 2 and our proposed system in Section 3. We present in Section 4 the background subtraction technique to highlight the moving objects in the scene. Section 5 is dedicated to outlining a robust approach for 3D localization of the moving objects. Results are detailed in Section 6. The conclusion is devoted to a discussion on the obtained results, and perspectives are provided.

2. 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. Any proposed system is not intended to replace the currently 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 localization 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 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. Generally, most level crossings are fitted with standard equipments such as lights, automatic full or half barriers, and notices. This equipment warns and prevents all users of the level crossing if a train is approaching the dangerous area.

3. Overview of the System

Our research aims at developing an Automatic Video-Surveillance (AVS) system using the passive stereo-vision principle. The proposed imaging system uses two cameras to detect and localize any kind of object lying on a railway level crossing. The system supervises and estimates automatically the critical situations by detecting objects in the hazardous zone defined as the crossing zone of a railway line by a road or path. The AVS system is used to monitor dynamic scenes where interactions take place among objects of interest (people or vehicles). After a classical image grabbing and digitizing step, this architecture is composed of the two following modules.

(i) Motion Detection Module. The first step consists of separating the motion regions from the background. It is performed using Independent Component Analysis (ICA) technique for high-quality motion detection. The color information is introduced in the ICA algorithm that models the background and the foreground as statistically independent signals in space and time. Although many relatively effective motion estimation methods exist, ICA is retained for two reasons: first, it is less sensitive to noise caused by the continuous environment changes over time, such as swaying branches, sensor noise, and illumination changes. Second, this method provides clear-cut separation of the objects from the background and can detect objects that remain motionless for a long period. Foreground extraction is performed separately on both cameras. The motion-detection step allows focusing on the areas of interest, in which 3D localization module is applied.

(ii) 3D Localization Module. This process applies a specific stereo matching algorithm to obtain a 3D localization of the detected objects. In order to deal with poor-quality images, a selective stereo-matching algorithm is developed and applied to the moving regions. First, a disparity map is computed for all moving pixels according to a dissimilarity function entitled Weighted Average Color Difference (WACD) [5]. An unsupervised classification technique is then applied to the initial set of matching pixels. This allows to automatically choose only wellmatched pixels. A pixel is considered as well-matched if its correspondant which is obtained thanks to a given stereo-matching algorithm, is the true correspondant. However, all true correspondants are given by the ground truth which allows to verify the accuracy of the applied matching algorithm. The classification is performed applying the confidence-measure technique detailed in [6]. It consists of evaluating the result of the likelihood function, based on

the "winner-take-all" strategy. However, the pixels constituting each object are then estimated applying a hierarchical belief-propagation technique detailed in Section 5.3.

4. Background Subtraction by Independent Component Analysis

4.1. Related Work. Real environments are much more complex than indoor environments and require advanced tools to deal, for instance, with sharp brightness variations. Another aspect that must be dealt with is the motion in the background, such as swaying branches, illumination changes, clouds, shadows, and sensor noise. Background subtraction is one of the motion detection methods introduced to extract the foreground objects from a reference background in an image sequence. In recent years, another set of techniques has emerged to cope with the problem of foreground estimation. The Independent Component Analysis (ICA) technique is getting much attention in video processing. It was introduced in the 1980s [7] in the context of neural network modeling. The purpose of ICA is to restore statistically independent source signals, given only observed output signals without knowing the mixing matrix of the sources. Zhang and Chen [8] have introduced the spatiotemporal ICA method to model a video sequence for background subtraction. Their scheme tries to extract a set of mutually independent components from a given mixture of two signals representing, respectively, a background and an image containing an arbitrary object. Recently, Tsai and Lai [9] have proposed an improved ICA scheme for background subtraction without background updating in indoor environment, but this method proves its effectiveness with a stationary monochrome camera. Their work is limited to an indoor environment with small environmental changes and only uses monochrome image sequences.

4.2. Motion Detection by Independent Component Analysis. ICA can be defined as a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It is a special case of blind-source separation. ICA defines a generative model for separating the observed multivariate data that are mixtures of unknown sources without any previous knowledge. Its aim is to find the source signals from observation data. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed to be non-Gaussian and mutually independent; they are called the independent components of the observed data. These independent components, also called sources or factors, can be found using ICA as shown in Figure 1.

An ICA algorithm can be seen as a convolution between two signals. The more the signals are similar, the smaller are the result values. In Figure 1, the intensity of pixels of the white lines on the road in the two images are very similar. The difference between these two signals gives a small value. The smaller the value, the darker the corresponding pixel. For background subtraction, the color ICA model outputs

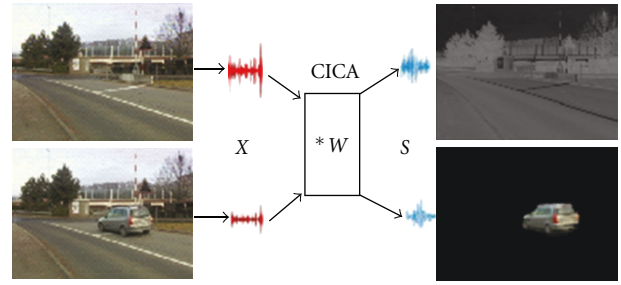


Figure 1: Principle of the background subtraction using independent component analysis. The input data is a combination of a reference signal, that is, background image (top left image), and a given signal, that is, an image from the sequence (bottom left image). ICA algorithm allows separating the input data into two independent signals: the one corresponds to the background model (top right image), the other corresponds to the estimated foreground without any details of the background (bottom right image).

three channels, each linked to a color component of the processed image: red, green, or blue. The channel with the highest signal/noise ratio is used to perform the motion-based segmentation process. The foreground segmentation is based on a threshold calculated from the histogram of the considered output channel. The threshold is estimated using the following procedure: each pixel in the output channel belongs to a class representing a color level. The color value corresponding to the class with the highest number of entries is taken as threshold (the entries of a class represent the number of pixels with a color corresponding to this class). Therefore, the foreground object can be easily extracted from the estimated source according to its Gaussian distribution.

In order to cope with outdoor environments, our framework aims at developing a novel Color-based Independent Component Analysis (CICA) model for motion detection in a color-image sequence. Initially, the ICA algorithm is performed in order to initialize the demixing matrix. The data matrix can be formed by both a random background image and another image containing the foreground objects, if any. The estimated source image corresponds to the modeled background and foreground images: one represents only the background source and the other contains the foreground object, without the detailed contents of the reference background. Figure 2 illustrates the synoptic of the algorithm. In our case, an image containing a foreground object is naturally independent, in time and space, from the background. The two-color images taken as an input to CICA are coupled in a matrix termed data matrix. CICA aims at separating the mixed signal into six separated signals: three channels per image, each channel representing a color component, for the foreground and background images.

The inverse of the mixing matrix, called de-mixing matrix, is estimated by the FastICA algorithm [10]. The estimated source images contain only the foreground object in a uniform region without the detailed contents of the background. The FastICA algorithm is based on a fixed-point iteration scheme maximizing non-Gaussianity as a

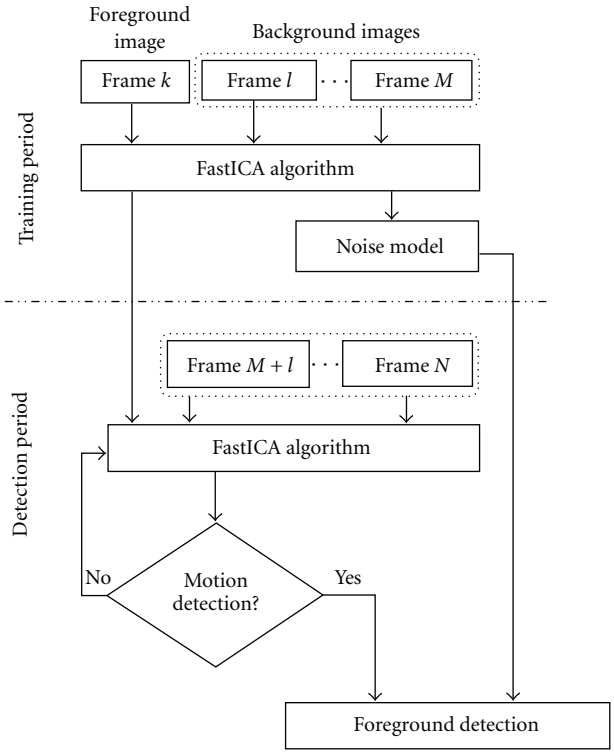


Figure 2: Global scheme of the CICA model for the object detection module.

measure of statistical independence. It attempts to find a set of independent components by estimating the maximum negentropy. The FastICA algorithm uses an approximation of the Newton method, tailored to the ICA problem, and provides fast convergence with little computation per iteration. In order to make this estimate, the algorithm iteratively searches for the weight set matrix of a neural network from a data set that properly separates the data signal mixtures into independent components. Let $E\{X \cdot X^T\}$ be the covariance matrix of the data matrix X. The nth iteration of the search loop makes an estimate of the nth weight vector. Note that an intuitive interpretation of the contrast functions is that they are measures of nonnormality. However, the estimated source signals are termed independent components. The iterative algorithm finds the direction for the weight matrix W maximizing the non-Gaussianity of the projection $W^T X$ for the data matrix X. The FastICA algorithm is described in Algorithm 1.

The observed mixture signal X is a linear combination of an unknown independent source signal S and an unknown de-mixing matrix W. It can be expressed as

$$S = W \cdot X. \tag{1}$$

Let the sample image $I^{x,y,k}$ be of size $(3,\ h \cdot w)$, where $x = 1 \cdots w$, $y = 1 \cdots h$, $k = \{r, g, b\}$, and h and w are the height and the width of the image, respectively. Each channel k is organized as a row vector of $l = (h \cdot w)$ elements. Let $X_{bg}^k = [x_{bg,1}^k, \ldots, x_{bg,l}^k]$ be the signal corresponding to channel k of the background image and $X_{fg}^k = [x_{fg,1}^k, \ldots, x_{fg,l}^k]$ the signal corresponding to channel k of the background containing an arbitrary object. Let data matrix X be of size $(6, h \cdot w)$ and defined as

$$X = \begin{pmatrix} x_{bg,1}^k & \cdots & x_{bg,l}^k \\ \vdots & \ddots & \vdots \\ x_{fg,1}^k & \cdots & x_{fg,l}^k \end{pmatrix}. \tag{2}$$

The de-mixing matrix W is of size 6×6. Each row is a weight vector that enables maximizing the independence between the different color channels. The foreground detection process is performed using CICA which is divided into two steps: noise modeling over a training period and moving object detection. An example of an estimated background and noise-models is shown in Figure 3. As mentioned previously, FastICA algorithm allows estimating independent signals from a mixture of these signals. The output signals corresponds to the difference between the input-signals. In our case, it corresponds to the variation of intensities of pixels between two consecutive images. In the noise modeling module, a set of background images is selected manually to initialize the CICA model. The input data signal to the CICA, that is, the data matrix, is formed by two consecutive frames providing, after performing the FastICA algorithm, a noise model. This operation is carried out during the training set. At the end of the training step, all elementary noise models are combined into only one model. For each incoming image from the sequence, a consecutive set of background images is selected automatically for noise model updating. An image is considered as a background if no motion is detected. The noise model will be used in the detection stage to denoise the output signal for a better detection. In the detection stage, the mixed signal is formed by two images: a background image and a scene image containing a foreground object. The noise model is updated during the detection period when the scene does not contain any moving object.

Indeed, this technique leads to the detection of any kind of objects such as pedestrians, cars, or arbitrary objects. Furthermore, one can highlight the advantages of this technique. First, the very small objects can be detected easily. Second, unlike other foreground detection techniques, CICA does not absorb a stationary object into the background. Therefore, the period during which an object is motionless does not affect the detection performances.

Unlike the existing methods based on a background subtraction scheme, the proposed CICA is less dependent on the background model. All swaying branches and illumination changes causes a uniform noise in the independent component obtained from the CICA which corresponds to the foreground signal. This uniform noise can easily be removed by filtering the foreground image with a Gaussian filter for instance. For a qualitative evaluation of the CICA model, a set of real-world image sequences are used. The most challenging dataset is the one containing swaying branches and clouds. This dataset is collected in a parking in Lausanne, switzerland. Figure 4 shows a white car in motion in difficult weather conditions. The CICA is compared to two

(1) Take a random initial guess for the weights associated with component $W_{i,0}$ of norm 1.
(2) Find $W_{i,n} = E\{x \cdot g(W_{i,n-1}^T \cdot x)\} - E\{g' \cdot W_{i,n-1}\} \cdot W_{i,n-1}$, where x is an observation from X an g is a contrast function (see discussion of the choice of g function in [10]).
(3) Divide w_n by its norm.
(4) If $|W_n^T \cdot W_{n-1}|$ is not close enough to one, then increment n and repeat starting at step (2).
(5) Repeat, starting at step (1), until all weights are found.

Algorithm 1: FastICA algorithm.

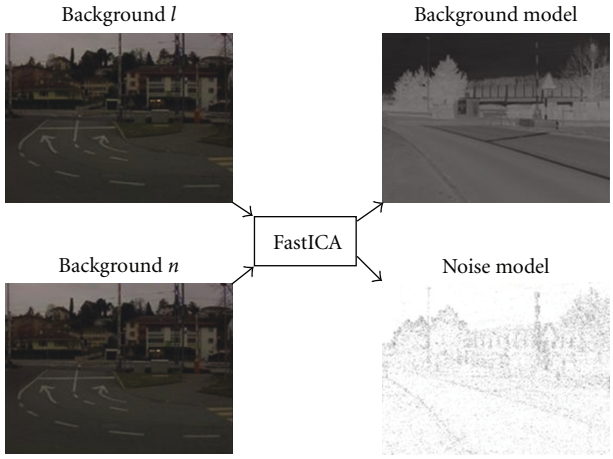


Figure 3: Noise modeling using ICA: (left) the n background images taken as input for ICA model; (right) the two outputs of the model are the background model and the noise model.

well-known methods which are Mixture of Gaussian [11] and Codebook [12]. In Figure 4(f), the foreground signal of our proposed method is not binarized in order to visually highlight the quality of results. For the other methods, (d) and (e), the choice of the threshold is a real challenge: the bigger the threshold, the fewer the foreground pixels and vice versa. In our case, the background of the estimated foreground is uniform. This allows to easily extract the real foreground.

An additional examples of foreground extraction are given by Figure 5. The first line corresponds to a pedestrian crossiong an LC in France in sunny weather. The second and the third line corresponds to a level crossing in "Pontet" in Switzerland in cloudy weather.

5. Stereo Matching for Robust 3D Localization

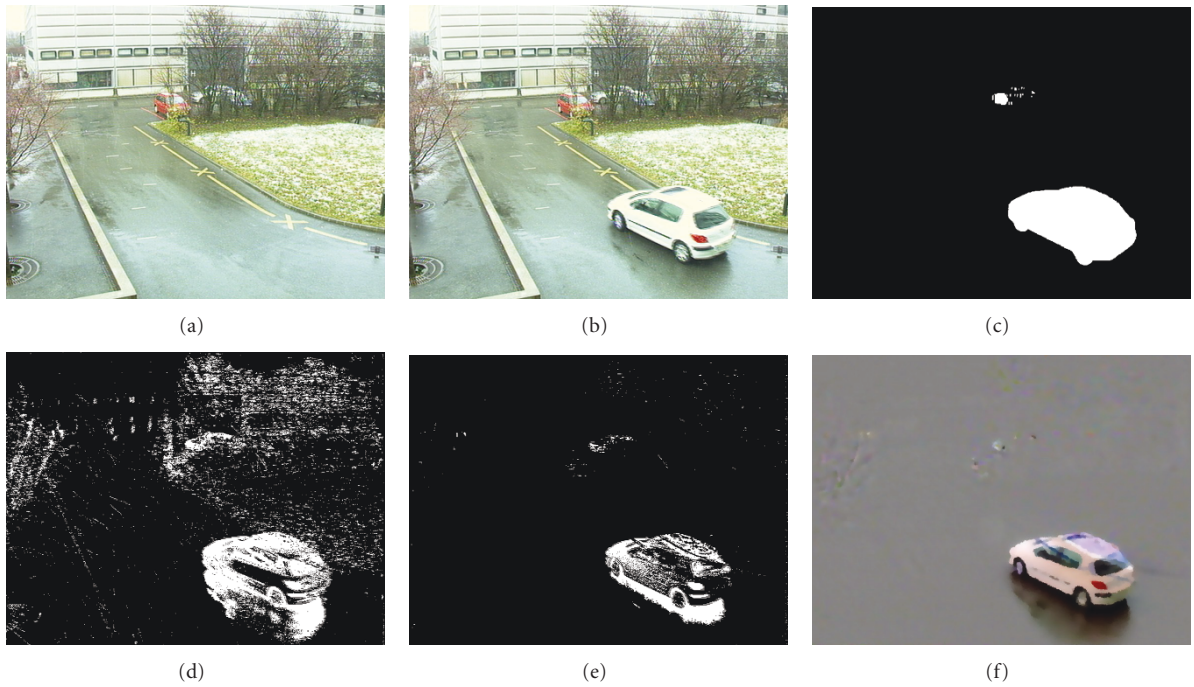
The two-frame stereo-matching approaches allow computing disparities and detecting occlusions, assuming that each pixel in the input image corresponds to a unique depth value. The stereo algorithm described in this section stems from the inference principle based on hierarchical belief propagation and energy minimization.

It takes into account the advantages of local methods for reducing the complexity of the belief-propagation method which leads to an improvement in the quality of results. A Hierarchical Belief Propagation (HBP) based on a confidence-measure technique is proposed: first, the data term (detailed in Section 5.1) is computed using Weighted Average Color Difference dissimilarity function (WACD) [5]. The obtained 3D volume allows initializing the belief-propagation graph by attributing a set of possible labels (i.e., disparities) for each node (i.e., pixels). The originality is to consider a subset of nodes among all the nodes to begin the inference algorithm. This subset is obtained thanks to a confidence measure computed at each node of a graph of connected pixels. Second, the propagation of messages between nodes is performed hierarchically from the nodes having the highest confidence measure to those having the lowest one. A message is a vector of parameters (e.g., possible disparities, (x, y) coordinates, etc.) that describes the state of a node. To begin with, the propagation is performed within each homogeneous color region and then passed from a region to another. The set of regions is obtained by a color-based segmentation using the meanshift method [13]. In level crossings, the motion constraint is also employed in the matching process in order to reduce both the matching error rate and the processing time. However, the 3D localization step concerns only the pixels in motion. A summary of our algorithm is given in Algorithm 2.

5.1. Global Energy Minimization. The global energy to minimize is composed of two terms: data cost and smoothness constraint, noted f and $\hat{f}$, respectively. The first one, f, allows to evaluate the local matching for each node by attributing a label l to a given node in a graph $\mathcal{G}$. The second term, $\hat{f}$, allows to evaluate the smoothness constraint by measuring how well label l fits pixel p given the observed data. The smoothness term is considered as the amount of difference between the disparity of neighboring pixels [14, 15]. This can be seen as the cost of assigning a label l' to a node during the inference step. the global energy minimization function can be formulated as

$$E(\mathcal{G}) = E_{l \in \mathcal{L}}(f) + E_{l' \in \mathcal{L}}\left(\hat{f}\right). \quad (3)$$

The minimization of this energy is performed iteratively by passing messages between all the neighboring nodes.



(a) (b) (c)

(d) (e) (f)

Figure 4: Example of foreground detection in a parking of the EPFL in Switzerland. (a) Background image that contains no object in motion. (b) Image from the sequence containing a moving white car with the presence of swaying branches. (c) Ground truth highlighting moving objects only. (d) Foreground objects obtained by Codebook technique. (e) Foreground objects obtained by MOG method. (f) Foreground objects obtained by our proposed CICA method.

These messages are updated at each iteration, until convergence. However, a node can be represented as a pixel having a vector of parameters such as, typically, its possible labels. Several studies [16–18] have proposed ways to improve the processing time of the inference process. However, reducing the complexity of the inference algorithm leads in most cases to reduced matching quality. Other algorithm variants can be derived from this basic model by introducing additional parameters in the message to be passed. A compromise must be found between the reliability and the computational cost. One of the important parameters is the spatiocolorimetric proximity between nodes [19].

(i) The data term we propose can be defined as a local evaluation of attributing a label l to a node. It is given by

$$E_{l\in\mathcal{L}}(f) = \sum_{p} \alpha\phi^{x\to x',y}(z_1), \tag{4}$$

where $\mathcal{L}$ is the set of all the possible disparity values for a pixel and $\phi^{x\to x',y}(z_1)$ is the cost obtained according to the WACD likelihood function of the couple of pixels $(p^{x,y}, p^{x',y})$. z_1 represents the first retained candidate having the lower cost. Parameter α is a fuzzy value within the $[0,1]$ interval. It allows computing a confidence measure for attributing a disparity value d to the pixel p. α is given by

$$\alpha = \begin{cases} \psi(p^{x\to x',y}) & \text{if } \psi(p^{x\to x',y}) \geqslant \rho, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\psi(p^{x\to x',y})$ is a confidence measure computed for each pair $(p^{x,y}, p^{x',y})$ of matched pixels and ρ is a confidence threshold. The way of computing the confidence technique is detailed in Section 5.2.

(ii) The smoothness term is used to ensure that neighboring pixels have similar disparities.

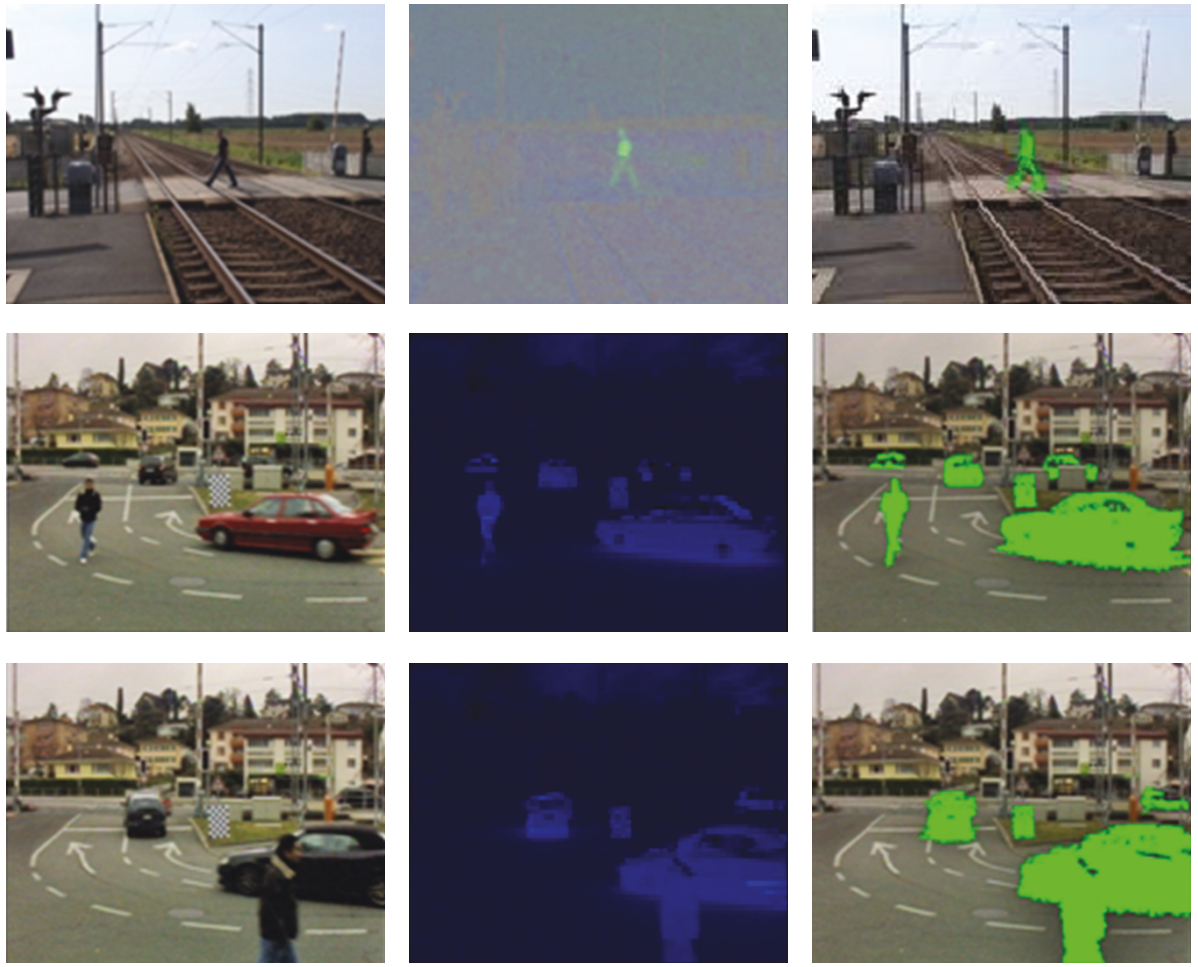
5.2. Confidence-Measure Computation. Using the WACD dissimilarity function allows initializing the set of labels. It represents a first estimate of the disparity map which contains matching errors. Then, each pair of pixels is evaluated using the confidence measure method described in [6]. The likelihood function used to initialize the disparity set is applied to each pixel of the image. Furthermore, for each matched pair of pixels a confidence measure is computed. It is termed $\psi(p_l^{x,y}, p_r^{x',y})$ which represents the level of certainty of considering a label l as the best label for pixel p. $p_l^{x,y}$ and $p_r^{x',y}$ represent the pixels in the left and right image, respectively, whose coordinates are (x,y) and (x',y). This confidence-measure function depends on several local parameters and is given by

$$\psi\left(p_l^{x,y}, p_r^{x',y}\right) = P\left(\frac{p_r^{x',y}}{p_l^{x,y}}, \rho, \min, \sigma, \omega\right). \tag{6}$$

The confidence measure with its parameters is given by

$$\psi\left(p_l^{x,y}, p_r^{x',y}\right) = \left(1 - \frac{\min}{\omega}\right)^{\tau^2\log(\sigma)}, \tag{7}$$

where



(a) (b) (c)

Figure 5: Several objects at level crossings. (a) Original images containing pedestrians, cars, and a stationary object. (b) Estimated signal corresponding to foreground signal. (c) Superimposition of the segmented moving objects into the original image.

(i) Best Correlation Score (min). The output of the dissimilarity function is a measure of 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 to be the actual correspondent.

(ii) Number of Potential Candidate Pixels (τ). This parameter represents the number of potential candidate pixels having similar scores. τ has a big influence because it reflects the behavior of the dissimilarity function. A high value of τ means that the first candidate pixel is located in a uniform color region of the frame. The lower the value of τ, the fewer the candidate pixels. If there are 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. A very small value of τ and a min score close to zero mean that the pixel to be matched probably belongs to a region of high color variation.

(iii) Disparity Variation of the τ Pixels (σ). A disparity value is obtained for each candidate pixel. For the τ potential candidate pixels, we compute the standard deviation σ of the τ disparity values. A small σ means that the τ candidate pixels are spatially 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 τ candidate pixels taken into account are situated in a uniform color region.

(iv) Gap Value (ω). This parameter represents the difference between the τth and $(\tau + 1)$th scores given with the dissimilarity function used. It is introduced to adjust the impact of the minimum score.

To ensure that function ψ has a value between 0 and 1, a few constraints are introduced. The min parameter

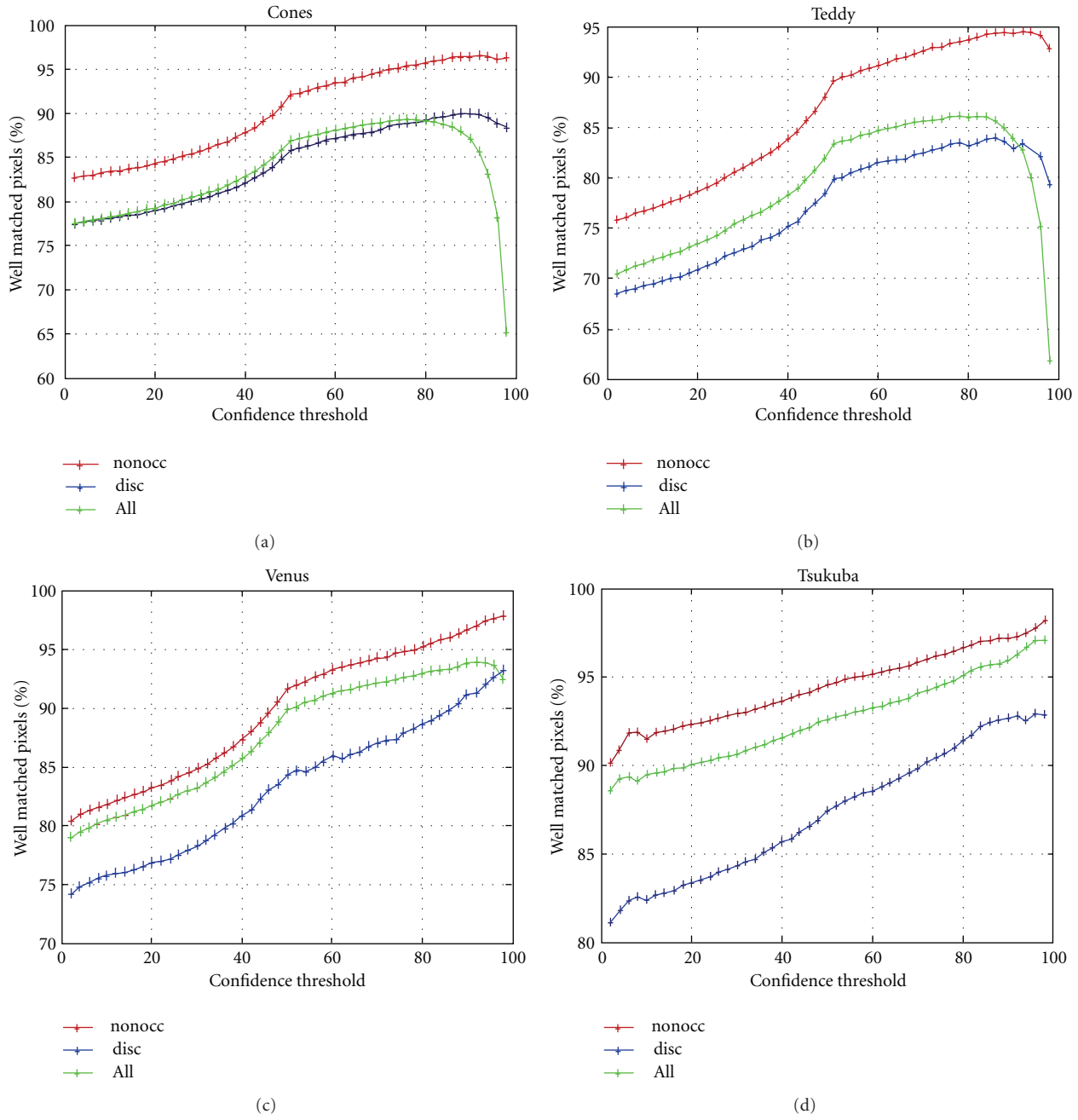


FIGURE 6: Percentage of well-matched pixels versus confidence-measure threshold in non-occluded regions (nonocc), all regions (all) and depth-discontinuity regions (disc) for (a) Cones, (b) Teddy, (c) Venus, and (d) Tsukuba datasets.

must not be higher than ω. If so, parameter ω is forced to min +1. Moreover, the $\log(\sigma)$ term is used instead of σ, so as to reduce the impact of high value of σ and obtain coherent confidence measures. The number τ of potential candidate pixels is deduced from the $\mathcal{L}$ scores obtained with the WACD likelihood function. The main idea is to detect major differences between successive scores. These major differences are called main gaps. Let ϕ denote a discrete function which represents all the scores given by the dissimilarity function in increasing order. We introduced a second function denoted η, which represents the average growth rate of the ϕ function. η 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

$$\eta\left(\phi^{x',y}\right) = \frac{\phi^{x',y}(z_m) - \phi^{x',y}(z_1)}{z_m - z_1} \quad m \in \mathcal{L}, \tag{8}$$

where $\phi^{x,y'}(z_m)$ is the likelihood cost obtained for the couple of pixels $(p^{x,y}, p^{x',y})$, z_m is the rank of the pixel $p^{x',y}$, $\eta(\phi^{x',y})$

(1) Initialize the data cost for nodes in the graph using the method in [5].
(2) Compute a Confidence Measure $\psi(p^{x \to x', y})$ for each node.
(3) Repeat steps (a), (b), (c) and d for each node
  (a) Select node (i.e., pixel) Node$_i$, i being the node number, having a data term lower than a confidence threshold ϱ.
  (b) Select the k-nearest neighbor nodes within a cubic 3D support window that have a $\psi(p^{x \to x', y})$ greater than ϱ.
  (c) Update the label of the current node.
  (d) Update the weight of the current node.
(4) Repeat step (3) until reaching minimal energy.

Algorithm 2: Hierarchical belief propagation.

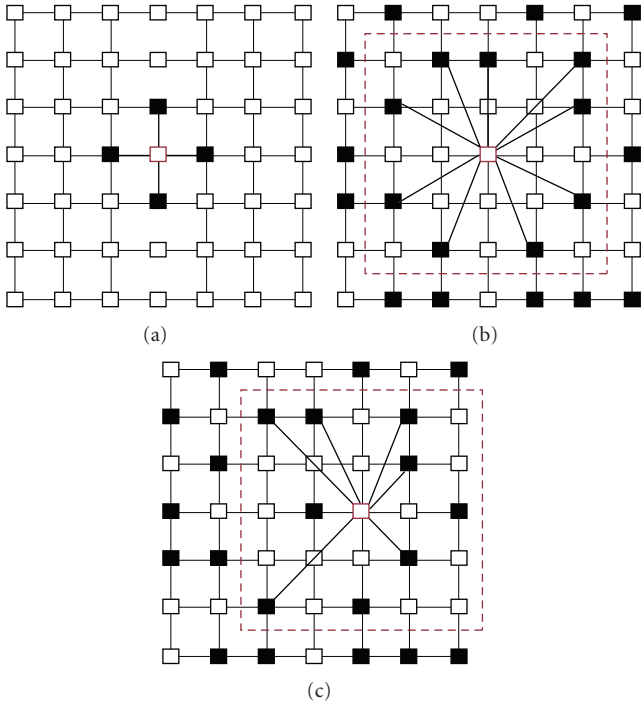


Figure 7: (a) Standard messages passing principle of the belief-propagation technique (b) iteration i of the proposed Hierarchcal Belief Propagation (HBP) based on the k-nearest neighbors having high confidence measure (c) iteration $i + 1$ of the HBP. The black square points denote pixels having a confidence measure higher than a fixed confidence threshold. The confidence of a given pixel can be upadted at the ith iteration and then it can be used in the $(i + 1)$th iteration.

is a discrete function that allows to highlight the large gaps between scores. It is materialized using

$$\xi\left(\phi^{x',y}\right) = \begin{cases} \dfrac{\nabla\eta^{x',y}}{m^2} & \text{if } \nabla\eta^{x',y} \geqslant 0, \\ -1 & \text{otherwise.} \end{cases} \tag{9}$$

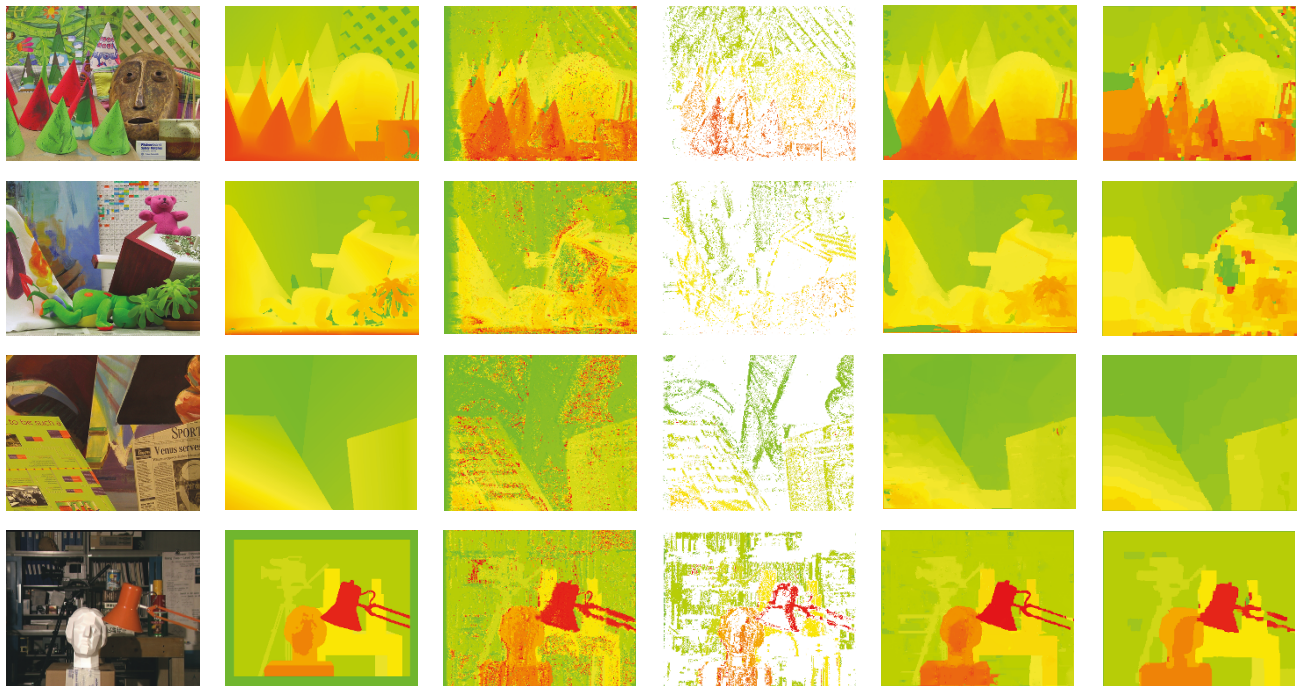
The previous function (7) is used to characterize the major scores and is applied only in the case where the gradient $\nabla\eta^{x',y}$ has a positive sign. We have introduced parameter m^2 in order to penalize the candidate pixels according to their rank. The number of candidate pixels is given by

$$\tau = \arg\max_{m} \xi\left(\phi^{x',y}\right). \tag{10}$$

Figure 6 shows the percentage of well-matched pixels depending on the confidence-measure parameter. The rate of well-matched pixels varies depending on the complexity of each pair of stereo images. A scene containing many textureless, uniform, or occluded regions decrease the number of well-matched pixels and vice versa. According to Figure 6, the number of well-matched pixels drops dramatically with a high confidence threshold for Cones and Teddy stereo images. In this case, the ambiguity of matching of most pixels is expressed by a small confidence measure. This is not the case for Venus and Tsukuba stereo images. For these latters, the ambiguity of matching decreases because of a high local color variation of most of pixels.

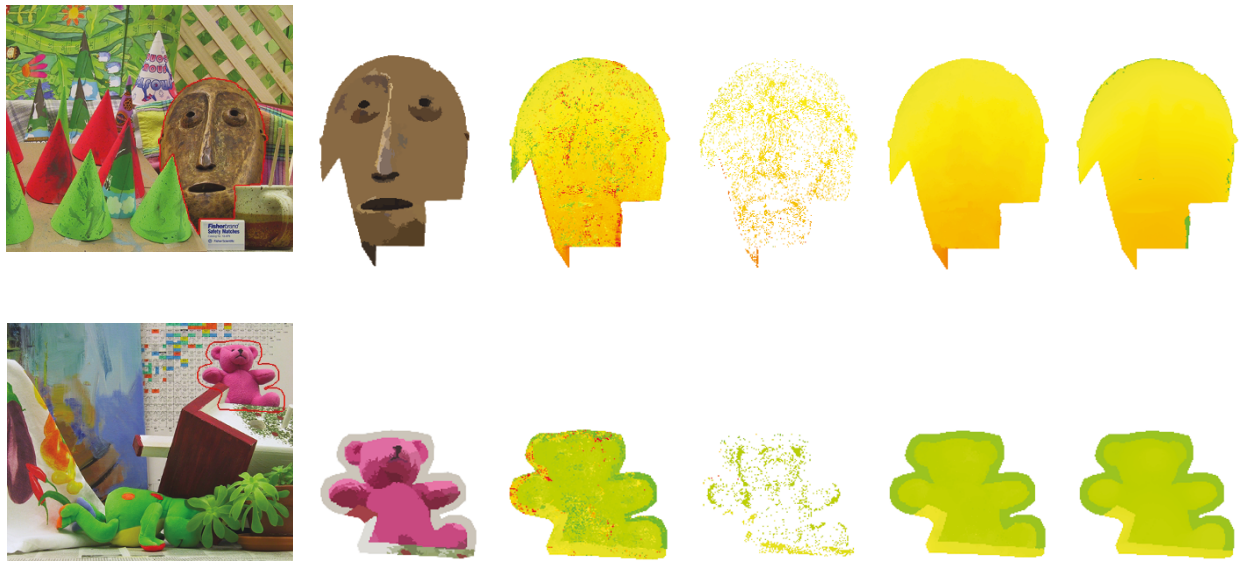
5.3. Hierarchical Belief Propagation for Disparity Enhancement. All the matched pixels can be modeled as a set of nodes in an undirected graph. Typically, the inference algorithm based on a belief-propagation method [20, 21] can be applied to achieve the optimal solution that corresponds to the best disparity set. A set of messages are iteratively transmitted from a node to its neighbors until convergence. Referring to this basic framework, all the nodes have the same weight, meaning that a message is passed from a node to all its neighbors. The main drawback is that several erroneous messages might be passed across the graph, leading to an increased number of iterations without guarantee of reaching the best solution. Several works have tried to improve the performances of the message passing step of the standard belief-propagation method. The proposed HBP technique allows both improving the quality of results and speeding up the inference step.

The messages are passed across the graph is as follows: we consider a cubic support window centered on the node to be updated. The latter receives messages only from the k-nearest neighbors obtained according to both their spatial proximity in the disparity space and their confidence measure. Figure 7 shows an example of message passing procedure. The central pixel to be matched (in red edges) is surrounded of a set of pixels having a high confidence (in



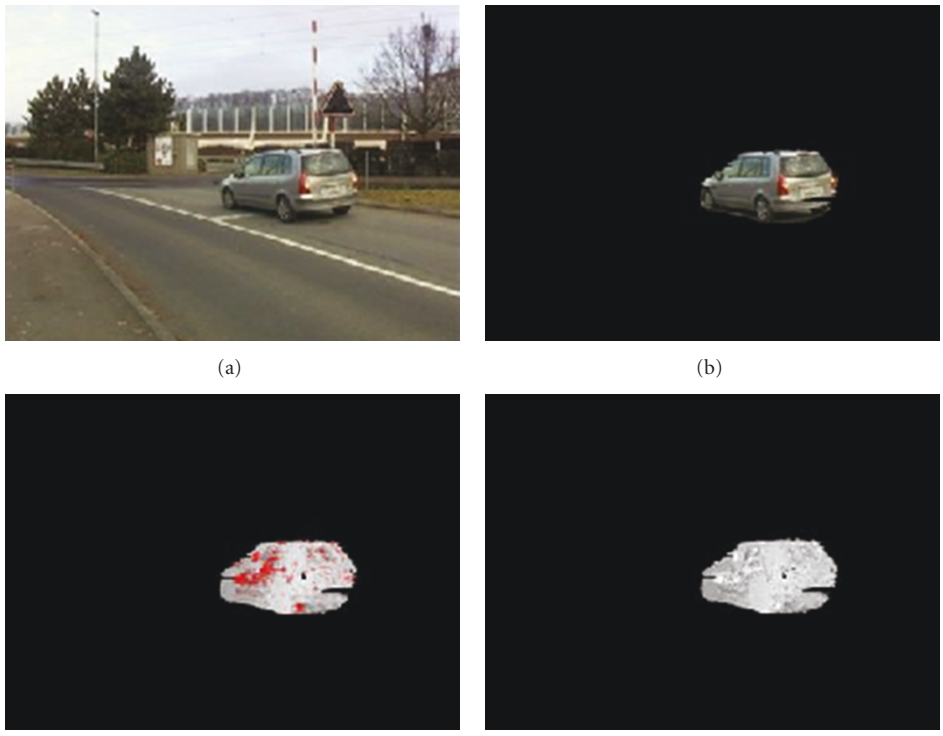
(a) (b) (c) (d) (e) (f)

Figure 8: Dense disparity map obtained with our algorithm compared to the max-product algorithm [14]. (a) Standard test sets. Top to bottom: Cones, Teddy, Venus and Tsukuba. (b) Ground truth for visual comparison. (c) Dense disparity map obtained applying the WACD likelihood function (d) Well-matched pixels having a confidence measure greater than 60%. (e) Dense disparity map obtained from our algorithm. (f) Dense disparity map obtained using the max-product formulation of the BP.



(a) (b) (c) (d) (e) (f)

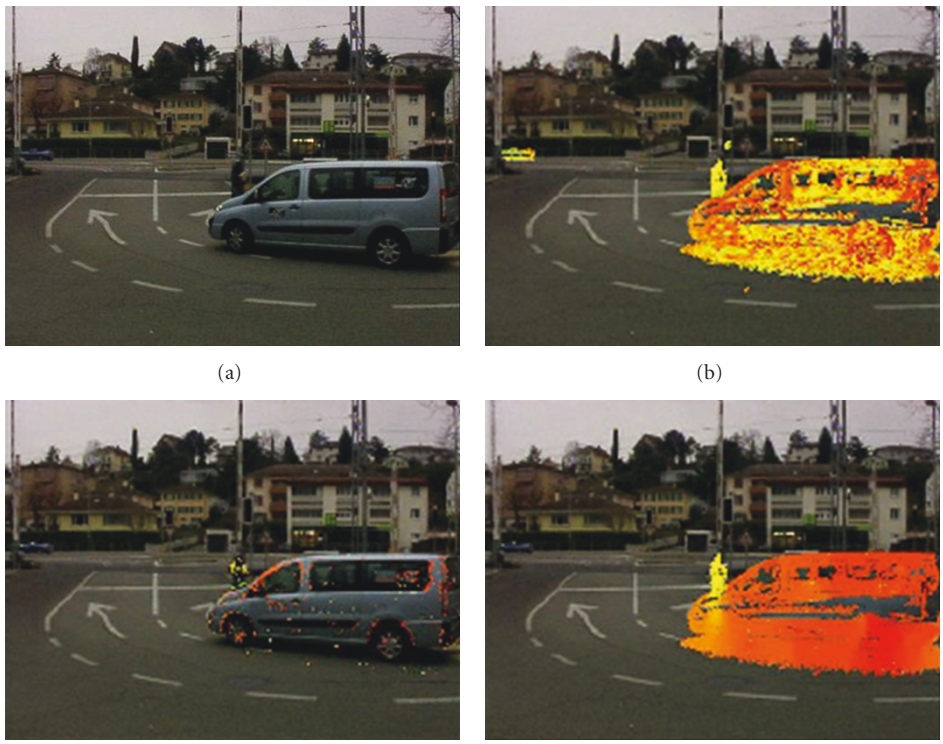
Figure 9: Different steps of our algorithm in different types of regions. (a) Left image for Cones and Teddy. (b) Segmented face and teddy bear extracted from the Cones and Teddy images, respectively, using mean shift. (c) Dense disparity map obtained using WACD. (d) Sparse disparity map corresponding to the well-matched pixels, with 60% confidence threshold. (e) Dense disparity map after performing the HBP. (f) Corresponding ground truth.



(a) (b)

(c) (d)

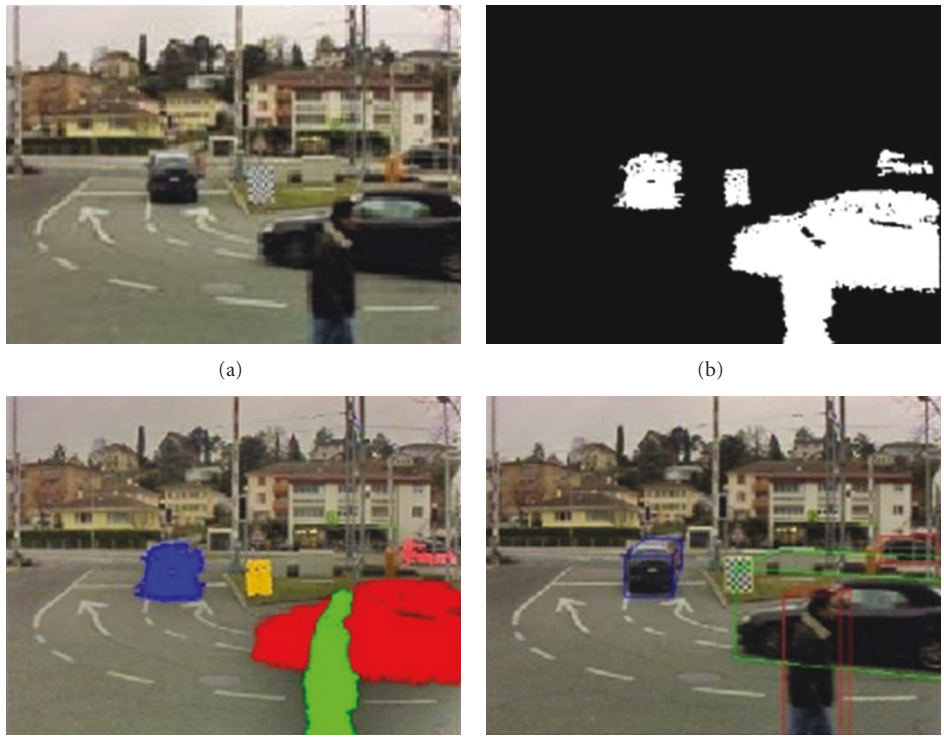
Figure 10: (a) Left-hand image. (b) Moving car extracted by CICA (c) Disparity map obtained by WACD likelihood function. (d) The red pixels are false matches. (f) Improved disparity map using confidence measure.



(a) (b)

(c) (d)

Figure 11: Results obtained using HBP based on confidence measure. (a) Left-hand image. (b) Sparse disparity map obtained using WACD in the moving regions. (c) Well-matched pixels estimated using the confidence measure with a threshold of 80%. (d) Final disparity map.



(a) (b) (c) (d)

Figure 12: Segmentation results using CICA. (a) Original image. (b) CICA segmentation result. (c) Object localization obtained by stereo matching. (d) Final 3D object localization.

black color) and others with low confidence (white color). The neighboring are grouped within a squared windows of size 5×5 (shown as discontinious red lines).However, the nodes having homogeneous disparity values and a high confidence measure are activated. Let Ω be the set of neighbor nodes which have a high confidence measure within the cubic support window of edge β. The subset of nodes, denoted by $\Omega^* \subset \Omega$, is chosen according to

$$\Omega^* : \left\{ p_i / \underbrace{\frac{1}{N}\sum_{k=1}^{N} \Delta(p_i, p_k)}_{d_i} < Z \right\}; \quad i \in \Omega,\ Z = \beta\sqrt{3}. \tag{11}$$

The disparity of the central node of the support window is updated according to

$$\hat{d} = \frac{1}{\sum_{i \in \Omega^*} d_i} \sum_{k=1}^{N} \mathrm{disp}_k \cdot d_k; \quad i, N \in \Omega^*, \tag{12}$$

where $\Delta(p_i, p_k)$ is the Euclidian distance between node p_i and node p_k in $\mathbb{R}^3$ and Z is the diagonal of the cubic support window of edge β. According to (11), noisy nodes, characterized by a high confidence measure and an outlying disparity value, are eliminated. This reduces the errors in the high level of the message passing step and enables to decrease significantly the number of iteration, which leads to reach the optimal solution quickly.

The main ideas of the HBP are detailed below.

(i) The confidence measure is used to assign a weight to each node in the graph. At each iteration, messages are passed hierarchically from nodes having a high confidence measure (i.e., high weight) to nodes having a low confidence measure (i.e., small weight). A high weight means a high certainty of the message to be passed. The weights of the nodes are updated after each iteration, so that a subset of nodes is activated to be able to send messages in the next iteration.

(ii) The propagation is first performed inside a consistent color region, and then passed to the neighboring regions. The set of regions is obtained by a color-based segmentation using the mean shift method [13].

(iii) In our framework, the messages are passed differently from the standard BP algorithm. Instead of considering the 4-connected nodes, the k-nearest neighboring nodes are considered. These k-nearest neighboring nodes belong to a cubic 3D support window. We assume that the labels of nodes vary smoothly within a 3D support window centered on the node to be updated.

Table 1: Parameter settings used in the experiments.

Mean shift			WACD	Hierarchical BP	
α_{ms}	β_{ms}	γ_{ms}	L_{sw}	knn	ϱ_{cm}
7	7	15	7	6	0.6

6. Experimental Results

The effectiveness of the proposed system is evaluated on both standard and real dataset. Each module is evaluated separately due to the unavailability of a conventional ground truth for motion estimation and 3D localization at a Level crossing.

The proposed stereo-matching algorithm is evaluated on the Middlebury stereo benchmark [22], using the Tsukuba, Venus, Teddy, and Cones standard datasets and on real-world datasets. The evaluation concerns nonoccluded regions (nonocc), all regions (all) and depth-discontinuity regions (disc). In the first step of our algorithm, the WACD likelihood function is performed on all the pixels. Applying the "winner-take-all" strategy, a label corresponding to the best estimated disparity is attributed to each pixel. The second step consists of selecting a subset of pixels according to their confidence measure. Indeed, the pixels having a low confidence measure generally belong to either occluded or textureless regions. However, the subset corresponding to the well-matched pixels is taken as the starting point of the hierarchical belief-propagation module. We begin by evaluating the selective approach of attributing a confidence-measure to each matched pair. Figure 6 shows the percentage of well-matched pixels depending on the confidence measure parameter. The higher the confidence measure, the greater the rate of well-matched pixels.

The experimental results are obtained by using the set of parameter settings summarized in Table 1. The likelihood function has a single parameter L_{sw}, the size of the support window. Then, the hierarchical belief propagation is performed by setting a confidence threshold ϱ_{cm} and the k-nearest neighbors (knn) before the inference step. The homogeneous color regions in which the propagation is performed are obtained by the mean shift color segmentation algorithm. It depends on the spatial bandwidth α_{ms}, the color bandwidth β_{ms}, and the minimum region size γ_{ms} parameters.

A first qualitative evaluation is proposed. Our framework compared to the efficient belief propagation detailed by Felzenszwalb [14]. The latter is a technique for speeding up the standard belief propagation combining three techniques: the linear time message updates, the bipartite graph message passing schedule by computing messages "in place", and the multigrid method for performing BP in a coarse-to-fine manner. The evaluation is performed on the Middlebury datasets. Figure 8 shows the disparity map obtained with the WACD likelihood function and the distribution of well-matched pixels obtained with a confidence threshold of 60%. The disparity of the remaining pixels is estimated by performing the HBP. This allows obtaining a dense disparity map.

Quantitatively, our method has been compared to several other methods from the literature. These methods are H-Cut [23], max-product [14], and PhaseBased [24]. Table 2 provides quantitative comparison results between these three methods and the proposed one. This table shows the percentage of pixels incorrectly matched for the nonoccluded pixels (nonocc), the discontinuity pixels (disc), and for all the matched pixels (all). More specifically, the proposed method is better for Tsukuba in "all" and "disc" pixels, in Venus for "disc" pixels and in Cones for "all" pixels.

Figure 9 illustrates an example of two objects extracted from the Cones and Teddy images, respectively. The face extracted from Cones corresponds to an nonoccluded region while the teddy bear corresponds to a depth discontinuity region. This proves that the propagation of disparities preserves the discontinuity between regions and gives a good accuracy in terms of matching pixels in the nonoccluded regions.

Several evaluations of our algorithms have been carried out in real-world situations. For this purpose, datasets have been acquired, composed of a hundred real scenarios of cars, pedestrians, objects, and so forth, crossing different LCs in France and Switzerland. These scenarios are either real events or played by actors in order to increase the number and variety of cases and allow evaluating in depth the accuracy of our obstacle detection system in terms of objects extraction and 3D localization (cf. Figure 10).

The different steps described in the previous sections are illustrated in Figure 10, showing a car crossing an LC in Lausanne (Switzerland). CICA is applied to the left-hand image. The segmentation results are used as motion constraints to the stereo matching process, yielding quite often several disparity values for each detected foreground point. The false matches corresponding to wrong disparity values are detected automatically using the confidence-measure technique. However, the final disparity map obtained for each object allows locating very precisely each object at the LC. The foreground extraction method based on CICA has already been evaluated in terms of Recall (95%) and Precision (98%), on a set of 300 images with manually elaborated ground truth.

The introduction of the confidence measure in the matching process improves the accuracy of the disparity of each segmented object. The disparity allows estimating the 3D position and spatial occupancy rate of each segmented object. The transformation of an image plane point $p = \{x, y\}$ into a 3D reference system point $P = \{X, Y, Z\}$ must be performed. The distance of an object point is calculated by triangulation, assuming parallel optical axes

$$Z = \frac{b \cdot f}{d}, \tag{13}$$

where Z is the depth, that is, the distance between the sensor camera and the object point along the Z axis, f is the focal length, that is, the distance between the lens and the sensor, supposed identical for both cameras, b is the baseline, that is, the distance separating the cameras, and d is the estimated disparity.

Table 2: Algorithm evaluation on the Middlebury dataset.

Algorithm	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
H-cut	2.85	4.86	14.4	1.73	3.14	20.2	10.7	19.5	25.8	5.46	15.6	15.7
Proposed	**4.87**	**5.04**	**8.47**	**3.42**	**3.99**	**10.5**	**17.5**	**20.8**	**28.0**	**7.46**	**12.5**	**13.3**
Max-product	1.88	3.78	10.1	1.31	2.34	15.7	24.6	32.4	34.7	21.2	28.5	30.1
Phase based	4.26	6.53	15.4	6.71	8.16	26.4	14.5	23.1	25.5	10.8	20.5	21.2

In complex images containing noise or complex structures, even the above methods are prone to false matching causing errors in the 3D localization. Figure 11 shows such a case, in which high sensor noise and the low-contrast texture of the car yield a very noisy disparity map (cf. Figure 11). The confidence measure is used to retain only the matched pairs having a high confidence value, so as to increase the robustness. Then, HBP is used to fill the homogenous regions to which those matched pairs belong. This allows eliminating the disparity noise caused by spurious matches, while retaining the disparity gaps between adjacent regions corresponding to objects at different distances. Occlusions are thus dealt with in a satisfactory manner for the application, instead of being mistakenly merged.

Figure 12 illustrates another important feature of the combined CICA plus stereo-matching approach, which allows in a first step to retain only the moving foreground of the scene (b), which is then separated into different objects using stereo vision (c, d). An interesting feature of the CICA approach for foreground segmentation is that, unlike other methods, it allows to retain an object that has moved into the scene but then has remained stationary, even for a long time. This is the case of the chessboard-like test card visible in Figure 12, as well as cars stopped at or on the LC—a very important feature in our application, since they are exactly what we need to detect.

7. Conclusions and Perspectives

In this paper, we have proposed a processing chain addressing safety at level crossings composed of a foreground extraction based on CICA followed by a robust 3D localization. The latter proves its effectiveness compared to stereo-matching algorithms found in the literature. The experimentations showed that the method is applicable to real-world scenes in level-crossing applications. The foreground extraction method based on CICA has already been evaluated in terms of Recall (95%) and Precision (98%) on a set of 300 images with manually elaborated ground truth. Real-world datasets have been shot at four different level crossings, including a hundred scenarios per level crossing under different illumination and weather conditions. The global chain including foreground extraction and 3D localization, still needs to be evaluated intensively on the above dataset. According to the experimentations, the localization of some objects may fail. However, the localization of one among sixty objects fails, this is due to the smaller number of pixels having confidence measure larger than a fixed threshold. The starting point of the belief-propagation process highly depends on the number and repartition of pixels, having hight confidence measure, inside an object. This drawback can be handled by introducing the temporal dependency in the belief-propagation process.

The main output of the proposed system is an accurate localization of any object in and around a level crossing. For safety purposes, the proposed system will be coupled with already existing devices at level crossings. For instance, the status of the traffic light and the barriers will be taken as input in our vision-based system. The level of such an alarm depends on the configuration of the different parameters. For instance, the presence of an obstacle in the crossing zone when the barriers are lowering is a dangerous situation, and the triggered alarm must be of high importance. A Preliminary Risk Analysis (PRA) seems to be an interesting way to categorize the level of alarms. In the frame of the French project entitled PANSafer, these different parameters will be studied. In particular, telecommunication systems will be used to inform road users on the status of the level crossing. Such informations could also be shared with the train driver and the control room. The communication tool and the nature of information to be transmitted are in study.

References

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

[2] http://pansafer.inrets.fr.

[3] 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.

[4] M. Ohta, "Level crossings obstacle detection system using stereo cameras," *Quarterly Report of Railway Technical Research Institute*, vol. 46, no. 2, pp. 110–117, 2005.

[5] 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 *Proceedings of the TELECOMA and 6th JFMMA*, pp. 206–209, Agadir, Maroc, 2009.

[6] N. Fakhfakh, L. Khoudour, M. El-Koursi, J. Jacot, and A. Dufaux, "A new selective confidence measure-based approach for stereo matching," in *Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, vol. 5711, pp. 184–191, Santiago, Chile, 2009.

[7] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural networks model," in *Proceeding of International Conference on Neural Networks for Computing*, pp. 206–211, Snowbird, Utah, USA, April 1986.

[8] X.-P. Zhang and Z. Chen, "An automated video object extraction system based on spatiotemporal independent component analysis and multiscale segmentation," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–22, 2006.

[9] 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.

[10] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[11] T. Zhen and M. Zhenjiang, "Fast background subtraction using improved GMM and graph cut," in *Proceedings of the 1st International Congress on Image and Signal Processing (CISP '08)*, vol. 4, pp. 181–185, Sanya, China, 2008.

[12] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real time foreground-background segmentation using a modified codebook model," *Journal of Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[13] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[14] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[15] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[16] M. Isard and J. MacCormick, "Dense motion and disparity estimationvia loopy belief propagation," in *Proceedings of the Asian Conference on Computer Vision*, pp. 32–41, Hyderabad, India, January 2006.

[17] X. Zhou and R. Wang, "Stereo matching based on color and disparity segmentation by belief propagation," *Optical Engineering*, vol. 46, no. 4, 2007.

[18] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nister, "Real-time global stereo matching using hierarchical belief propagation," in *Proceedings of the Brith Machine Vision Conference (BMVC '06)*, pp. 989–998, September 2006.

[19] H. Trinh, "Efficient stereo algorithm using multiscale belief prpagationon segmented images," in *Proceedings of the Brith Machine Vision Conference (BMVC '08)*, 2008.

[20] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2009.

[21] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1458–1465, San Francisco, Calif, USA, June 2010.

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

[23] D. Miyazaki, Y. Matsushita, and K. Ikeuchi, "Interactive shadow removal from a single image using hierarchical graph cut," in *Proceedings of the Asian Conference on Computer Vision (ACCV '09)*, 2009.

[24] S. El-Etriby, A. Al-Hamadi, and B. Michaelis, "Desnse stereo correspondance with slanted surface using phase-based algorithm," in *Proceedings of the IEEE International Symposium on Indistrual Electronics*, 2007.