Detection and Recognition of Urban Road Markings Using Images

Philippe Foucher, Yazid Sebsadji, Jean-Philippe Tarel, Pierre Charbonnier and Philippe Nicolle

Abstract— **While road lane markings detection was extensively studied, in particular for intelligent vehicle applications, the detection and recognition of all kind of marking such as arrows, crosswalks, zebras, words, pictograms, continuous and discontinuous lane markings was drastically less studied. However, it has many potential applications in the design of advanced driver assistance systems, as well as for asset management along itineraries. An algorithm is proposed which is based on the following processing steps: marking pixel extraction, detection using connected components before inverse perspective mapping and recognition based on the comparison with a single pattern or with repetitive rectangular patterns. The proposed algorithm is able to detect and recognize repetitive markings (such as crosswalks) as well as single patterns (such as arrows). We believe that the proposed algorithm can be extended easily to solve the problem of the identification of all types of markings.**

I. INTRODUCTION

Road lane marking detection from onboard video sensors was extensively studied over the last two decades, see˜[11] for an overview. Indeed, many Advanced Driver Assistance Systems (ADAS) rely on the estimation of the vehicle's position with respect to the road along time. This generally implies the local detection of, at least, the left and right lane markings. Examples of ADAS in which lane marking detection is of major importance are road lane keeping systems, lane departure warning systems, and collision avoidance systems. While most systems require the determination of the lateral position and orientation of the vehicle in its lane, sometimes with road curvature, only a few of them incorporate *road marking recognition*. This explains why this problem was drastically less studied than lane markings detection. However, solving both problems, i.e. detection *and* recognition, may be beneficial to many ADAS. For example, the knowledge of line modulation is necessary for overtaking or lane change assistance systems. Such systems also require the detection and recognition of left, straight and right arrows. Indeed, the detection and recognition of all kinds of markings should provide future ADAS with a more comprehensive knowledge of the vehicle's environment. Detection and recognition algorithms also provide important information for road managers. Using a dedicated inspection vehicle, it allows to quickly update the knowledge about road makings along an itinerary, for asset management.

It must be noticed that marking detection and recognition are subject to several difficulties: presence of other white objects, low contrast in shadows, dust and erasing, occlusion by vehicles and other obstacles. In road marking recognition, it is important to distinguish between two classes of markings. The first one is the class of regular lane markings, which are rectangular shapes of different sizes. The second one is the class of specific markings and contains more diverse shapes, such as arrows, crosswalks, zebras, writings, and pictographs. Our final goal is to propose a road marking recognition algorithm which is generic in the sense that the method can be used to recognize all kinds of lane and specific road markings. However, the proposed work was developed in the context of the French iTowns project, where important landmarks such as arrows and crosswalks are used to register aerial images with terrestrial images acquired from an observation vehicle in urban environments. As a consequence, the proposed algorithm focuses on arrows and crosswalks in urban environment.

The paper is organized as follows. We first propose a short review of related works in Sect.˜II. Then in Sect.˜III, we describe the steps of the proposed algorithm: marking pixel extraction, detection using connected components before inverse perspective mapping, recognition based on comparison with a single pattern or with repetitive rectangular patterns. The experimental setup is presented in Sect.˜IV. In Sect.˜V, the experimental results and the evaluation results are discussed for crosswalks and 5 types of arrows.

II. RELATED WORK

Several specific marking detection and recognition algorithms were proposed focusing on pedestrian crossings. For instance, in˜[16] the detection and recognition method is based on an accumulation technique, namely the Hough transform; in˜[2], graphical models are used and, in˜[21], projective invariants are exploited. More recently, in˜[18], a crosswalk detection algorithm using edge extraction, Fourier transform and Hough transform has been proposed. In [19], [6], the algorithm consists in filtering images in frequency domain and detecting hatched marking by a RANSAC approach. It is important to notice that these algorithms are dedicated to pedestrian crossings and can not be easily extended to other kinds of specific markings, except zebra crossing and hatched marking. A method based on stereovision was also proposed in˜[20] which is dedicated to rectangular road markings and thus to crosswalks. Note that we here assume a single calibrated camera. Few works are focused

Philippe Foucher, Yazid Sebsadji and Pierre Charbonnier are with LRPC Strasbourg (ERA 27 IFSTTAR), 11 Rue Jean Mentelin, BP 9, F-67200 Strasbourg, France, {philippe.foucher, pierre.charbonnier}@developpement-durable.gouv.fr and y.sebsadji@gint.fr

Jean-Philippe Tarel is with Univ. Paris-Est, IFSTTAR, 58 Bd Lefebvre, F-75015 Paris, France, jean-philippe.tarel@ifsttar.fr

Philippe Nicolle is with IFSTTAR, route de Bouaye/CS4 44344 Bouguenais, France, philippe.nicolle@ifsttar.fr

on arrows. In˜[5], arrows are detected by region-based segmentation, followed by connected components extraction and geometric selection on the obtained regions. Other features are proposed to classify arrows according to their types: shape signature and Fourier descriptors in˜[15], geometric moments in˜[8], or a combination of geometric moments, angles and histograms in˜[7]. In˜[13], a generative model considering appearance variations is applied on standard road markings to create learning images. The recognition is then proposed by computing distance between by the L largest eigenvalues of candidate image and reference images. In these contributions, it may be noticed that an Inverse Perspective Mapping (IPM)˜[10], [1] is systematically applied to the original images beforehand, unlike the following two papers˜: a model fitting algorithm is performed on graylevel perspectives images to compare candidate contours with prototypes objects, encoded as arc splines˜[9]. In [3], the authors propose to recognize road marking in intersection by a 3-step algorithm˜: segmentation with gaussian mixture, 3D reconstruction and classification by thresholding six features. In summary, each previously cited work focuses on a single class of specific road marking while the challenge is the detection and the recognition of the diverse classes of road markings.

III. Road marking recognition

The road marking recognition system we propose consists of two steps: extraction of marking elements and identification of resulting connected components as crosswalks or directional arrows.

A. Marking Element Extraction

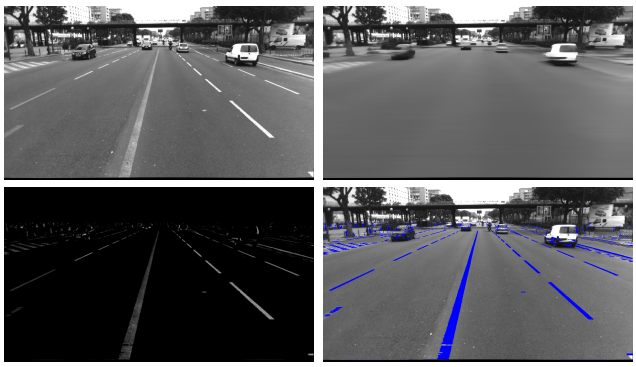

Fig.˜1.˜˜ 1^{st} line: original image and median filtered image. 2^{nd} line: original image after local threshold $T_G = 20$ using the filtered image and the extraction result (extracted marking pixels are in blue).

The classification of the image pixels in two classes: “marking” and “non-marking” was previously studied in depth in [22], [17]. In [22], we started from an experimental comparison between several road marking elements extraction algorithms, on a set of 116 ground-truth images named the ROMA database (available at www.lcpc.fr/en/produits/ride/). The ground-truth database is made of images taken from a vehicle along with a hand-made segmentation of the marking pixels in each image. This database was extended with more than one hundred extra images taken in various weather and visibility conditions (also available at www.lcpc.fr/en/produits/ride/ under the name ROMA2). The ROMA and ROMA2 databases contain mainly regular lane markings and only a few specific markings. As a consequence, we built another ground-truth database, named MiTowns, with high-resolution images (of size 1920×1080) taken in the urban area of Paris where many specific markings can be observed. The three databases, ROMA, ROMA2 and MiTowns were used in˜[17] to compare several extraction algorithms, for both lane markings and specific road markings extraction. We then proposed several improvements, for instance to take into account the fact that specific markings are most often wider than lane markings. At the time of the publication of˜[17], the MiTowns database was made of 47 images, and now it contains 80 images[1].

In summary, the best extraction algorithm we found˜[17] processes the image line-by-line independently, seeking for segments of marking pixels. In the processing, the horizontal sizes are modulated for each line to take into account the perspective effect the road is subject to when observed by a frontal camera. The basis idea of the algorithm relies on the fact that the intensity of a marking pixel is higher than the intensity of pavement pixels belonging to its neighborhood. The three steps are illustrated in Fig.˜1. The first step of the extraction algorithm consists in filtering the original image to remove all the white objects with an horizontal size lower than the maximum horizontal size of markings. The second step is a local thresholding that classifies as “marking” every pixel in the original image with a value higher than the value of the corresponding pixel in the filtered image plus a threshold T_G. Other pixels are classified as “non-marking”. The result of the second step is hence a binary image. Last, the third step consists in removing the horizontal segments of marking that are too small to be considered as marking elements.

Several different filters might be used in the first step. Taking a median filter leads to the so-called Median Local Threshold (MLT) algorithm˜[4], [12]. However, we found that the best results [17] are obtained by using the value of the 43rd percentile, instead of the 50th as in the median. The algorithm is thus called the 43rd Percentile Local Threshold (43rd PLT) extraction algorithm.

The ROC curves established on the 80 images of the MiTowns database is shown in Fig.˜2. This figure confirms that the 43rd PLT extraction algorithm performs better than the MLT extraction algorithm, as first shown in˜[17]. It must be underlined that after optimization, the complexity of the 43rd PLT extraction algorithm is a linear function of the number of pixels in the image and does not depend on the size of the filter.

When the input image is in RGB, it was proposed in [22] to process independently each color channel and then to fuse the three resulting binary images using a *min* operator. We

[1] Soon available at www.lcpc.fr/en/produits/ride/

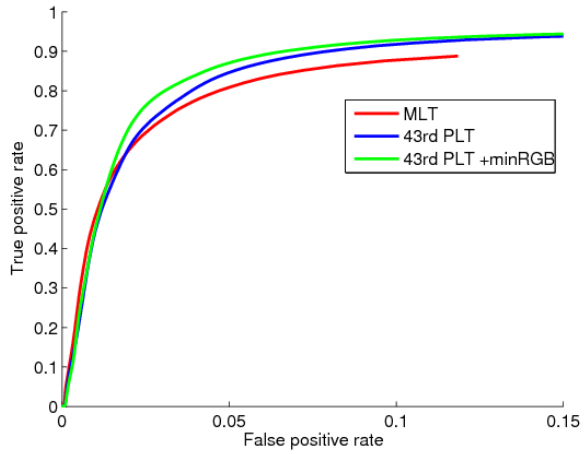

Fig.˜2.˜˜ ROC curve obtained using Median Local Threshold extraction algorithm on the color image (MLT), using the 43rd percentile Local Threshold extraction algorithm on the color image (43rd PLT) and using the 43rd PLT on the min over the color components (43rd PLT+minRGB).

found that it is better to merge the color channels before extracting road marking elements, by taking the minimum over the three color components. This divides the processing time by a factor of three. For instance, the average processing time on a dual-core processor 2.4 GHz, for an image of size 1920×1080, is 100ms. Moreover, as shown by the ROC curve in Fig.˜2, the use of the minimum over the color components leads to better results. This relatively important improvement can be explained by a better robustness to road specular highlights. In Fig.˜3, road specular highlights can be seen between the zebra strips, which result in a loss of contrast between road and zebra in several color channels. Therefore, the zebra is badly extracted using the MLT and 43rd PLT extraction algorithms on the original color image as illustrated by the second and third images of Fig.˜3. The fourth image shows the minimum over the three color components, where we can notice how the contrast between the road and the zebra is improved. The 43rd PLT extraction algorithm is thus able to better segment the zebra strips as illustrated in the last image of Fig.˜3.

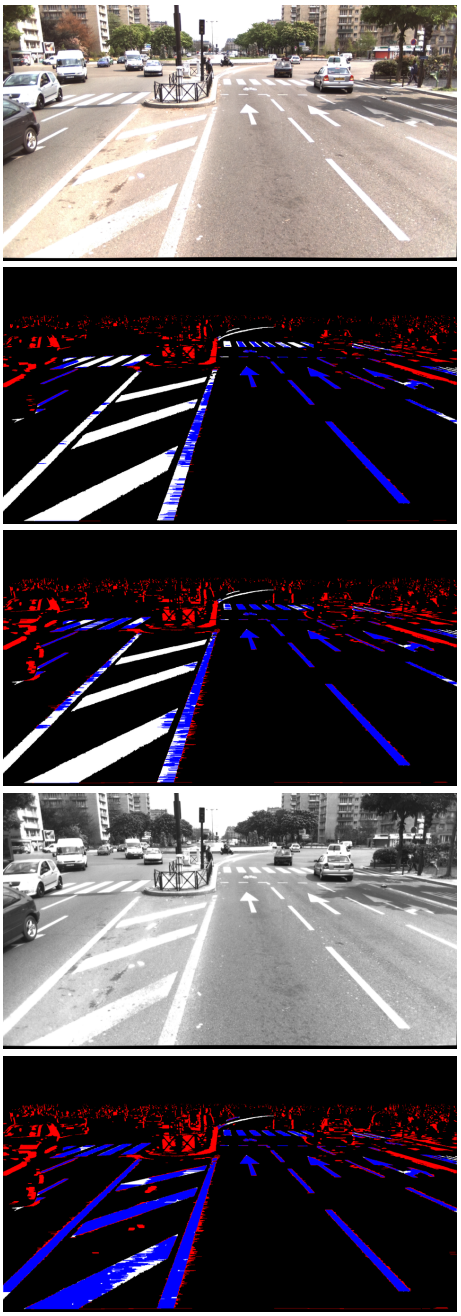

Fig.˜3.˜˜ From top to bottom: original color image; extraction results of the MLT algorithm and of the 43$^{\text{rd}}$ PLT algorithm on the color image; min over the color components of the original image; extraction results of the 43$^{\text{rd}}$ PLT algorithm on the min image. Red: False Positives, blue: True Positives, white: False Positive. The same threshold value is used ($T_G = 20$). Notice how the contrast is improved between road and zebra in the fourth image.

B. Identification

The identification step consists in detecting objects as connected components and then in analyzing their shape. The recognition algorithm is applied after inverse perspective mapping (IPM) to avoid perspective deformations and scale changes˜[10], [1] (see Fig.˜4). In the IPM image, the connected components are first selected according to several geometric parameters to reduce the number of candidates:

- the dimensions of the major axis L and the minor axis l of the connected components must lie within the ranges $[\sigma^L_{min}, \sigma^L_{max}]$ and $[\sigma^l_{min}, \sigma^l_{max}]$. The thresholds are fixed from theoretical values given by regulatory norms.
- the rectangularity criterion is defined as $(L_{mean} \times l_{mean})/(L_{max} \times l_{max})$ where L_{mean} and l_{mean} are the mean dimensions of the connected component in the direction of major and minor axis, and L_{max} and l_{max} are the maximum dimensions in the direction of major and minor axis (see fig.˜5).

The threshold values of the geometric filtering step are gathered in Tab.˜I. The recognition of crosswalks and arrows are then processed independently. Indeed, crosswalks belong to the class of markings with repetitive patterns and arrows

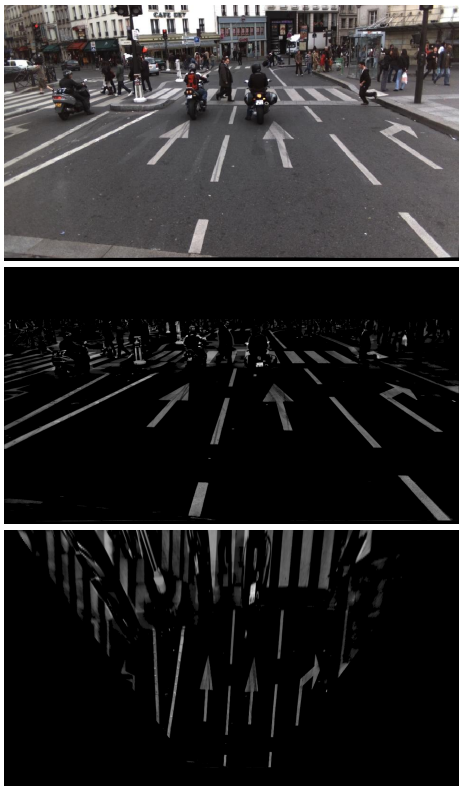

Fig.˜4.˜˜ original image (top) ; road marking extraction (middle) ; Inverse perspective mapping (bottom).

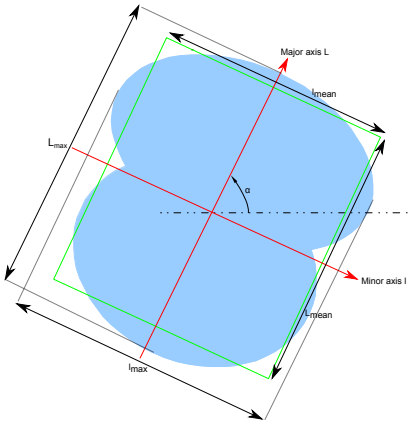

Fig.˜5.˜˜ Rectangularity criterion used during the geometric filtering of the connected components.

TABLE˜I

THRESHOLD VALUES USED THE THE GEOMETRIC FILTERING OF THE CONNECTED COMPONENTS.

	Crosswalks	Arrows
σ^L_{min} (in pixels)	135	80
σ^L_{max} (in pixels)	405	320
σ^l_{min} (in pixels)	20	30
σ^l_{max} (in pixels)	60	-
Rectangularity	upper than 0.7	lower than 0.7

are markings with a single pattern.

1) Crosswalks recognition: The crosswalks recognition is performed by considering the location of connected components related to each other. Two connected components, at least, belong to the same crosswalk if their orientations α are similar and if their centroids are located on the same minor axis and are separated by a maximum distance $9 \times l_{mean}$. This value is voluntary fairly high to take into account partial occlusions of the object of interest. These partial occlusions are most often due to masking by a vehicle. An exemple of this situation is presented in Fig.˜6-1st row.

2) Arrows recognition: The arrows recognition is performed by comparing extracted connected components to the 5 different types of directional arrows. Actually, several instances of a same arrow with partial occlusion perturbations are considered and 63 models of arrows are represented in the model data set. The candidate image, that corresponds to the bounding box of resulting connected component and the model image are normalized to the size 120×120. The models have been learned with their rotated position within $-15°$ et $+15°$ related to major axis in the image plane. The similarity criterion firstly considers the binary Hamming distance between the model image and the candidate image. If the Euclidean distance is lower than a threshold σ_h, the connected component is recognized with the type of closest model. Otherwise, a second similarity criterion is computed between the histograms of the closest model and the histograms of the candidate. The two projection histograms are defined by the sum of object pixels along major axis L and minor axis l. Note that the histograms can be computed by the discrete Radon transform for the angles α and $\alpha + \pi/2$. The similarity is then computed as the intersection between the histogram of the model and the histogram of the candidate. The identification is confirmed if the similarity values are upper than σ_L for major axis and σ_l for minor axis. At the moment, the values σ_h, σ_L, σ_l have been chosen experimentally on a few examples.

IV. EVALUATION DATA SET

The ground-truth database contains 280 images of size 1920×1080 of road scenes in urban environment for a total amount of 165 crosswalks and 151 arrows. The crosswalks are of variable sizes and seen at different distances. The five types of arrows appear in the database and the arrows are seen at various distances. An image may contain none or several objects of interest. The reference is a hand made set of bounding boxes around every crosswalk and arrow. Every bounding box is labeled as crosswalk or arrow.

A. Evaluation metrics

A detection is a bounding box in the images. A reference bounding box is considered as correctly detected when it exists a detection bounding box close enough with the correct label (crosswalks or arrow). We denote (x_r, y_r) the center of the reference bounding box, (w_r, h_r) its width and height and α_r its orientation. The same notations with subscript d are used for the detected bounding box. A candidate is considered as true positive (TP) if˜:

- the distance between (x_r, y_r) and (x_d, y_d) is lower or equal to $0.3max(w_r, h_r)$,
- the height h_d is lower or equal to $0.45h_r$,
- the absolute difference of orientation $|\alpha_d - \alpha_r|$ is lower or equal to 20 degrees.

Given a reference bounding box, if the previous conditions are not verified for any detection bounding box, it is a false negative (FN). Given a detection bounding box, if the previous rule is not verified for any reference bounding box, it is a false positive (FP). Usually, in the receiver operating characteristic (ROC) curve, the true positive rate (TPR) is plotted versus the false positive rate (FPR). The true positive rate is defined as $TPR = \frac{TP}{TP+FN}$. The false positive rate is defined as $FPR = \frac{FP}{FP+TN}$. The difficulty in computing FPR is to estimate the number of true negative TN which is very large and thus which artificially bias towards zero the FPR. Rather than using the false positive rate, it is more meaningful to use the false positive rate per image (FPPI) which is defined as the ratio of FP over the number of images in the database.

V. Experimental results

The first experiment focuses on the optimal value of the threshold T_G. Indeed, the efficiency of the extraction algorithm has a strong impact on the performances of the whole system. The value of T_G which maximizes the true positive rate is chosen as optimal. On a sub-database containing 94 images with 61 crosswalks and 46 arrows, the detection and recognition algorithm is applied and the TPR is computed for each integer value T_G in the range $[0, 255]$. The best TPR is obtained for $T_G = 20$ with 95% of true detection for crosswalks (3 crosswalks are not detected) and 87% of true detection for arrows (6 arrows are not detected). A careful examination of the results shows that non-detected markings are usually occluded by object of the complex environment (see Fig.˜6-3^{rd} row). Other non-detected markings are highly erased or very dirty (see Fig.˜6-4^{th} row). Nevertheless, the algorithm is able to detect worn-out crosswalks and arrows as shown in Fig.˜6-3^{rd} and 5^{th} rows, and is robust to back-lights as illustrated in Fig.˜6-2^{nd} row. It may be noticed that only 6 false alarms have been detected in the 94 images ($FPPI = 0.064$). In the second experiment, the algorithm is tested on the 186 remaining images in which 104 crosswalks and 105 directional arrows appear. The TPR is 90% for crosswalks (10 non-detected makings) and 78% for directional arrows (23 non-detected markings). Like in the first experiment, the missed objects are damaged (erased or dirty) or occluded markings. It is important to notice that 100% of detected arrows are correctly identified among the five types of arrows. Concerning the false alarms rate, 3 crosswalks and 7 arrows have been wrongly detected and identified in the 186 images. As a consequence, the false positive rate per image is fairly low: $FPPI = 0.053$. It may be pointed out that the TPR is always lower for arrows than for crosswalks. A possible explanation is that crosswalks are wider with repetitive patterns and thus the detection and recognition algorithm is more robust to occlusion. Indeed, while one or several patterns may be occluded, there is, in general, enough non-occluded patterns for a correct detection and recognition. The case of arrows detection is different because it is a single pattern which can be directly affected by occlusion and extraction errors. From the results we obtained, it seems that the first way to increase the performances of the algorithm may be to complete the model data set with additional reference images by adding other perturbations such as additive noise. A second way may be to improve the marking extraction step by introducing, for instance, edge information. The figure˜7 shows an example of marking extraction using both 43^{rd} PLT method and edge detection.

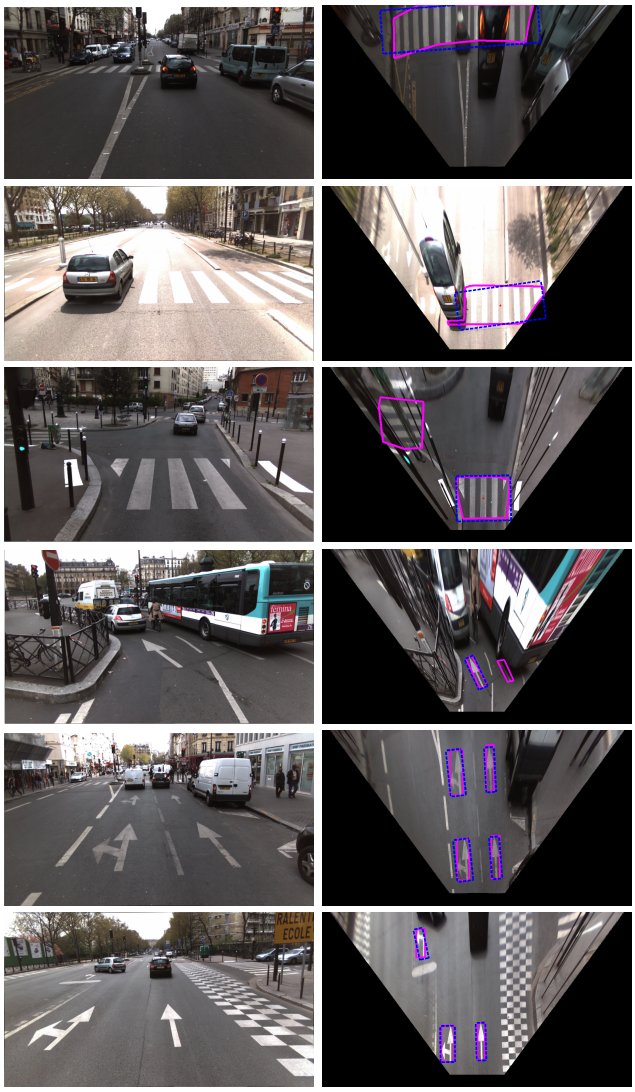

Fig.˜6.˜˜ 1^{st} col.: original images ; 2^{nd} col.: marking recognition. Ground-truth bounding boxes are in magenta and algorithm results are in blue.

VI. Conclusions and future works

In this paper, we have proposed an algorithm for detecting and identifying crosswalks and arrows in urban images. We have tested this algorithm on a data base containing real-world, high-resolution images with a non-negligible amount

Fig.˜7.˜˜ From top to bottom : original image ; road marking extraction using 43^{rd} PLT ; road marking extraction using 43^{rd} PLT (in gray level) and edge information (in yellow).

of difficult cases. The results show that 90% of crosswalks and 78% of arrows have been detected. We believe that the method proposed in this paper may be adapted to identify more generic markings such as zebras, words or pictograms. At short-term, we should investigate the extension of the set of models, the use of edge informations in the extraction step and the combination of two extractions as presented in˜[14]. At mid-term, other investigations should be leaded to optimize the recognition algorithm. More specifically, the number of thresholds involved should be reduced and the remaining ones should be set by a systematical analysis on a training database. These improvements should be confirmed by further evaluations on a larger test database and will hopefully increase the true detection rate.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Stéphane Sadowski in building the reference databases. This work is partly funded by the ANR (French National Research Agency) within iTowns-MDCO project.

REFERENCES

[1] M.˜Bertozzi, A.˜Broggi, and A.˜Fascioli. Stereo inverse perspective mapping: theory and applications. *Image and vision computing*, pages 585–590, 1998.

[2] J.˜Coughlan and H.˜Shen. A fast algorithm for finding crosswalks using figure-ground segmentation. In *2nd Workshop on Applications of Computer Vision, in conjunction with ECCV*, volume˜5, 2006.

[3] R.˜Danescu and S.˜Nedevschi. Detection and classification of painted road objects for intersection assistance applications. In *Proceedings of IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 433–438, Funchal, Portugal, 2010.

[4] F.˜Diebolt. *Reconnaissance des marquages routiers*. Phd dissertation, Université Louis Pasteur, Strasbourg I, France, Dec. 1996. (in French).

[5] U.˜Franke, D.˜Gavrila, S.˜Gorzig, F.˜Lindner, F.˜Paetzold, and C.˜Wohler. Autonomous driving goes downtown. *IEEE Intelligent Systems and their Applications*, 13(6):40–48, 1998.

[6] T.˜Gavrilovic, J.˜Ninot, and L.˜Smadja. Frequency filtering and connected components characterization for zebra-crossing and hatched markings detection. In *IAPRS, Vol.XXXVIII, Part3A*, pages 43–48, Saint-Mand, France, Sept. 2010.

[7] A.˜Kheyrollahi and T.˜Breckon. Real-time road marking recogniton using a feature driven approach. *Machine Vision and Applications*, pages 1–11, 2010.

[8] Y.˜Li, K.˜He, and P.˜Jia. Road markers recogntion based on shape information. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 117–122, Istanbul, Turquie, 2007.

[9] G.˜Maier, S.˜Pangerl, and A.˜Schindler. Real-time detection and classification of arrow markings using curve-based prototype fitting. In *Proceedings of IEEE Intelligent Vehicles Symposium (IV'2011)*, pages 442–447, Baden-Baden, Allemagne, June 2011.

[10] H.˜Mallot, H.˜Bulthoff, J.˜Little, and S.˜Bohrer. Inverse perpsective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.

[11] J.˜McCall and M.˜Trivedi. Video based lane estimation and tracking for driver assistance: Survey, system, and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):20–37, 2006.

[12] J.˜Ninot, J.-P. Tarel, T.˜Gavrilovic, L.˜Smadja, and K.˜Heggarty. Amélioration de la reconnaissance des marquages routiers par l'optimisation d'algorithmes d'extraction. In *Proceedings of colloque COGIST'09*, St Quay Portrieux, France, 2009. (in French).

[13] M.˜Noda, T.˜Takahashi, D.˜Deguchi, I.˜Ide, H.˜Murase, Y.˜Kojima, and T.˜Naito. Recognition of road markings from in-vehcile camera images by a generative learning method. In *Proceedings of IAPR conference on Machine Vision applications*, pages 514–517, Yokohama, Japan, 2009.

[14] E.˜Pollard, D.˜Gruyer, J.-P. Tarel, S.-S. Ieng, and A.˜Cord. Road marking extraction with combination strategy and comparative evaluation on synthetic and actual images. In *Proceedings of IEEE Conference on Intelligent Transportation Systems*, Washington DC, USA, 2011.

[15] J.˜Rebut, A.˜Bensrhair, and G.˜Toulminet. Image segmentation and pattern recognition for road marking analysis. In *Proceedigns of IEEE International Symposium on Industrial Electronics*, pages 727–732, Ajaccio, France, 2010.

[16] S.˜Se and M.˜Brady. Road feature detection and estimation. *Machine Vision and Applications*, 14(3):157–165, 2003.

[17] Y.˜Sebsadji, J.-P. Tarel, P.˜Foucher, and P.˜Charbonnier. Robust road marking extraction in urban environments using stereo images. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 394–400, San Diego, USA, 2010.

[18] S.˜Sichelschmidt, A.˜HAselhoff, and A.˜Kummert. Pedestrian crossing detecting as a part of an urban pedestrian safety system. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 840–844, San Diego, USA, 2010.

[19] L.˜Smadja, J.˜Ninot, and T.˜Gavrilovic. Global environment interpretation from a new Mobile Mapping System. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 941–948, San Diego, USA, 2010.

[20] B.˜Soheilian, N.˜Paparoditis, D.˜Boldo, and J.˜Rudant. Automatic 3d extraction of rectangular roadmarks with centimeter accuracy from stereo-pairs of a ground-based mobile mapping system. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 22–27, Padua, Italy, 2007.

[21] M.˜Uddin and T.˜Shioyama. Robust zebra-crossing detection using bipolarity and projective invariant. In *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, volume˜2, 2005.

[22] T.˜Veit, J.˜P. Tarel, P.˜Nicolle, and P.˜Charbonnier. Evaluation of road marking feature extraction. In *Proceedings of IEEE Conference on Intelligent Transportation Systems*, pages 174–181, Beijing, China, 2008.