

A Novel Two-Stage Algorithm for Non-Parametric Cast Shadow Recognition

Martin Roser and Philip Lenz

Institute for Measurement and Control Systems
Karlsruhe Institute of Technology (KIT)
D-76131 Karlsruhe, Germany
Email: {roser, lenz}@kit.edu

Abstract—Environment perception and scene understanding is an important issue for modern driver assistance systems. However, adverse weather situations and disadvantageous illumination conditions like cast shadows have a negative effect on the proper operation of these systems.

In this paper, we propose a novel approach for cast shadow recognition in monoscopic color images. In a first step, shadow edge candidates are extracted evaluating binarized channels in the color-opponent and perceptually uniform CIE $L^*a^*b^*$ space. False detections are rejected in a second verification step, using SVM classification and a combination of meaningful color features. We introduce a non-parametric representation for complex shadow edge geometries that enables utilizing shadow edge information for improving downstream vision-based driver assistance systems. A quantitative evaluation of the classification performance as well as results on multiple real-world traffic scenes show a reliable cast shadow recognition with only a few false detections.

I. INTRODUCTION

Proper and robust operation in arbitrary situations is a security-relevant prerequisite to vision-based driver assistance systems (DAS). Whereas state-of-the-art computer vision algorithms work reliably in good weather situations, adverse illumination conditions can cause severe problems. In particular, cast shadows induce color and illumination discontinuities in the image that can lead to unwanted behavior of image segmentation, object detection as well as correspondence search and hence have a severe influence on DAS like collision avoidance, lane keeping or intersection assistance [1], [8], [16], [19]. The decomposition of a single image in illumination and reflectance information is a classic and ill-posed problem in the field of computer vision: the number of unknowns is twice the number of equations. In the last decades, many assumptions have been proposed towards solving this ambiguity: Wu and Tang [20] propose a semi-automatic, Bayesian framework that removes shadows from single images by relying on a rough user input. Gevers et al. [9] introduce a rule-based reflectance classifier that distinguishes between different edge types, like highlight edges and material edges. However, a clear distinction between shadow edges and reflectance edges could not be drawn. Weiss [17] separates reflectance edges from illumination edges for a given sequence of images, assuming a static scene (constant reflectance edges) which is subject to illumination changes (moving shadow edges). It follows that a temporal median, applied to all gradient maps

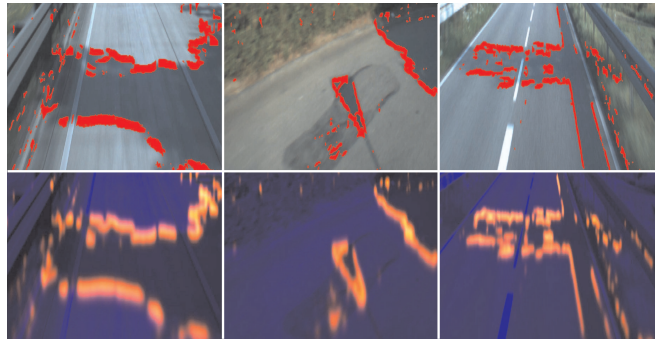


Fig. 1. Non-parametric cast shadow detection. The top row depicts verified shadow edge points. A non-parametric representation in terms of a shadow edge point density is given in the bottom row.

of the image sequence, preserves all stationary reflectance edges but suppresses shadow edges because the illumination direction is changing over the course of a day. This method is well appropriate for static surveillance cameras with temporal moving sun light, but in the field of automotive applications the problem statement is defined vice-versa: the illumination is assumed constant whereas the egomotion of the vehicle leads to a change of perspective, and therefore to a spatial shift of reflectance edges and shadow edges at the same time. Finlayson et al. [5] detect and remove cast shadows from monoscopic color images, using a color and intensity invariant *intrinsic image* [7] and lightness recovery techniques. Whereas the results look rather promising, their approach demands for a precise color calibration under different daylight conditions. Since the camera sensitivity may change as a function of time and temperature, obtaining such a calibration for automotive applications might be an expensive and time-consuming challenge. In [6] they further propose an approach for shadow removal by iteratively minimizing the entropy in the illumination invariant image. This allows for removing shadows in images acquired from uncalibrated cameras, but at the expense of an increasing computational time. Lalonde et al. [14] estimate the most likely illumination direction in outdoor scenes from a combination of weak image features. Amongst others, they extract shadow edges by thresholding the color channels in the CIELAB color space [15]. A large amount of missed

detections and clutter does not affect their application, since they use the shadow edge gradients in a voting scheme, only for estimating the predominant illumination direction.

In this paper, we present a two-stage approach for detecting cast shadows in single color images. First, we find shadow edge candidates by utilizing properties of the color opponent CIE $L^*a^*b^*$ color space (Section II). In a verification step, we propose a combination of meaningful point, local and global features as explained in Section III and use a machine learning technique to decide about the shadow edge candidate (Section IV). We finally discuss a possibility of processing shadow edge information in a non-parametric sense (Section V), making the information available for all downstream vision-based DAS. We conclude the paper with a summary and give an overview of future work.

II. SHADOW EDGE CANDIDATES

As observed in [12], color opponent and perceptually uniform spaces qualify as the most suitable representation for the purpose of shadow edge classification in outdoor scenes. Particular the CIE $L^*a^*b^*$ color space provides a clear edge distinction: strong shadow gradients are visible in the L^* channel but not in the a^* channel, whereas reflectance edges exhibit high gradient magnitudes in both channels (see Fig. 2).

As depicted in Fig. 2, we convert the RGB image I to the CIE $L^*a^*b^*$ color space and compute the horizontal and vertical image derivatives in the L^* and a^* channels. Then, we perform noise suppression on the gradient magnitudes ∇L^* , ∇a^* as well as adaptive thresholding

$$\begin{aligned}\nabla L_T^{*\sigma} &= \{\nabla L^* \otimes g_\sigma > T_L\} \\ \nabla a_T^{*\sigma} &= \{\nabla a^* \otimes g_\sigma > T_a\},\end{aligned}$$

where the Gaussian kernel g_σ is of size 5×5 pixels and its standard deviation is set to $\sigma = 1$. T_L and T_a are multiples of the mean values in a 64×64 pixel neighborhood in $L^{*\sigma}$ and $a^{*\sigma}$, respectively. A combination gives us all shadow edge candidates s :

$$s = \{x \in I : (\nabla L_T^{*\sigma}(x) = 1) \cap (\nabla a_T^{*\sigma} = 0)\}.$$

Note, that we need to consider high sensor noise, particularly in adverse illumination conditions and additionally waive any color calibration for the imaging sensor. For these reasons, the so far proposed low-level image processing approach produces many false detections which makes a verification step as discussed in Section IV necessary. Consequently, since we do not want to loose shadow points in the first detection step, the thresholds are set to low values ($T_L = 1.2\bar{w}$, $T_a = 1.7\bar{w}$ where \bar{w} are the local mean values) which results in a high recall (few false negatives) but a low precision (many false positives).

III. FEATURE EXTRACTION

In order to decide about the shadow edge candidates and reject as many false positives as possible, we extract a combination of different kinds of features from the RGB, HSV and CIE $L^*a^*b^*$ color space.

- **Point features:** Point features give information about the candidate point directly, without considering its neighborhood. Especially when dealing with thresholds in the beforehand processed detection algorithm, it is important to regard the actual pixel values for classification purposes. We consider the gradient magnitude values ∇L^* and ∇a^* as well as ∇H , ∇S , ∇V as the most important point features and include them into our descriptor vector.
- **Local features:** Local features add important, contextual information to the "isolated" shadow edge point. In particular, RGB color histogram features are promising, since they are robust to small displacements of the shadow edge and account for the typical blue tint of shadows on the road surface. Additionally, they are centered on the shadow edge and hence are expected to exhibit a bi-modal shape: one cluster with high values, representing the sunlit part of the road, and one cluster with low values for the shadowed region. In contrast, reflectance edges on structured scenes are expected to have multi-modal histograms. Altogether, we extract 45 features from a local 9×9 pixel neighborhood, which are the maximum value, mean value and standard deviation of the L^* , a^* and each channel in the HSV color space, respectively, as well as RGB color histograms with 10 bins for each color channel.
- **Global features:** These features may account for global illumination conditions and scene content. We extract the global maximum and mean value of the L^* , a^* and each HSV channel, respectively.

All feature values are extracted and combined into one extended descriptor vector, so we get the vector $\mathbf{v} = (v_1, \dots, v_n)$ with $n = 60$ scalar elements describing each shadow edge candidate.

IV. SHADOW EDGE VERIFICATION

We use the extended descriptor vector \mathbf{v} as described in the last section to either verify or reject the shadow edge candidate s . In our case, the verification step can be regarded as a binary classification problem into the classes $C = \{\text{shadow edge, reflectance edge}\}$, *i.e.* the challenge is finding some function f that maps from descriptor space \mathcal{D} into the classes C with $c = f(\mathbf{v})$, where $f : \mathcal{D} \rightarrow C$.

For a descriptor space with a small number of dimensions, such a function f can be designed by hand, whereas for high-dimensional descriptor spaces as discussed here ($\mathcal{D} = \mathbb{R}^{60}$), this might not be possible. However, a machine learning framework can be used to find such a function from training examples. Numerous methods have been proposed [2], [4] using techniques like k-Nearest-Neighbor, Decision Trees, Neural Networks and Support Vector Machines (SVM).

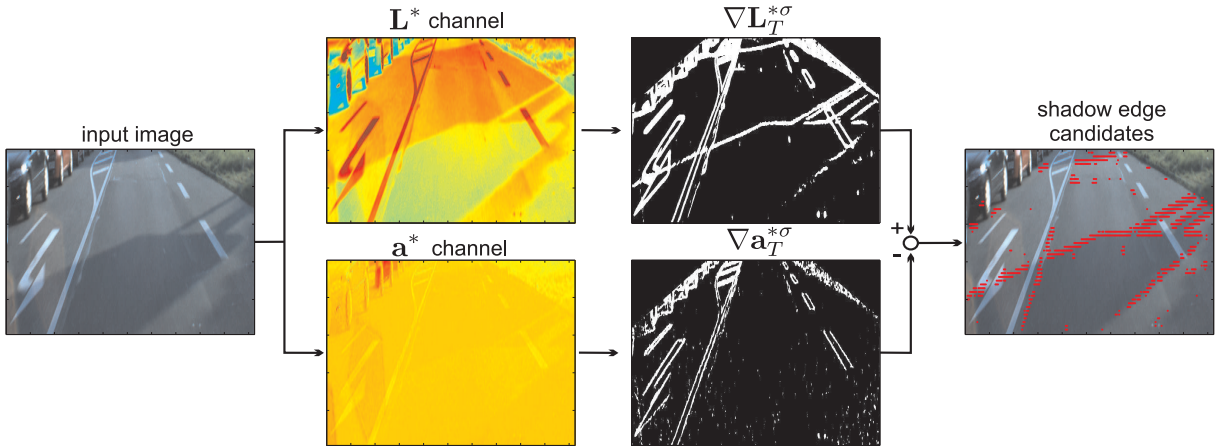


Fig. 2. **Shadow edge detection.** The RGB input image is converted to CIE $L^*a^*b^*$ color space. A combination of the thresholded gradient magnitudes gives a set of shadow edge candidates (red dots).

As SVMs are simple, fast, and powerful, we decided to use them for classification. In principle, a linear SVM generates a hyperplane in the descriptor space \mathcal{D} and classifies descriptors by calculating on which side of the hyperplane the descriptor vector (=point) lies. Mathematically, the hyperplane is represented by its normal vector \mathbf{w} with offset b , then for a given descriptor \mathbf{v} a score is calculated by $d = \mathbf{w}^T \mathbf{v} - b$ and the final decision is obtained by $d \geq 0$.

The hyperplane parameters \mathbf{w} and b are optimized in the learning stage to separate the two classes as far as possible. After training, the weights vector \mathbf{w} can be evaluated to get the significance of single features for the classification outcome. Large values correspond to discriminant features, whereas small values indicate weak features.

One of the advantages of SVMs is that kernel methods can be incorporated. With these kernels, non-linear decision boundaries can be found. We tested two very common kernels, linear and RBF (Radial Basis Functions), with the result that an RBF kernel outperforms the linear one. The RBF parameters that separate the two classes best are found to be $C = 1$ and $\gamma = 0.1$.

V. POST PROCESSING

In many indoor and outdoor applications with defined geometries, an appropriate low-dimensional parametrization of the detected shadow points, e.g. in terms of shadow lines [13], is desirable. However, we observed a large variety of different shadow shapes in common traffic scenarios which makes the choice of a capable parametric representation challenging. For this reason, we aim to derive a non-parametric but continuous shadow point density map.

We approximate the shadow point density on the road plane with a Gaussian convolution kernel. As depicted in Fig. 3(a), the actual kernel size for each image row can be determined, assuming planarity in front of the vehicle. A kernel size of $0.3 \times 0.3\text{m}$ with standard deviation $\sigma = 0.08\text{m}$ yields satisfying results (see Fig. 3(d)). In terms of computational time, the most convenient choice is a uniformly weighted box kernel as depicted in Fig. 3(c), which can

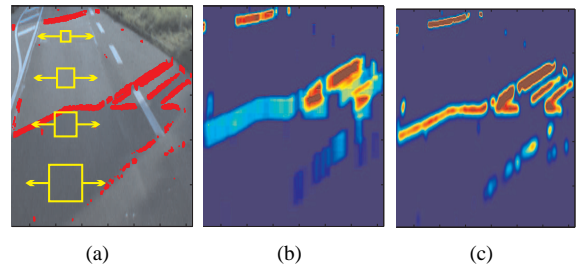


Fig. 3. **Non-parametric representation of cast shadows.** (a) illustrates the convolution of shadow edge points with a convolution kernel of variable size. (b) and (c) depict the shadow point density maps, achieved with a box and a Gaussian kernel, respectively.

be implemented very efficiently, using integral images [3]. However, multiple speedup efforts exist even for Gaussian kernels, e.g. by approximating a Gaussian kernel with the n^{th} repeated convolution of a box filter [10] or by a linear combinations of weighted integral images [11], [18].

Note, that the discussed density map can directly be used for modeling the probability of shadow edges and hence plays a major role when propagating the acquired information to downstream DAS.

VI. RESULTS

For all experiments, we randomly extracted images from a large database, containing more than 27 hours of expressway scenes, rural streets and urban environments under various weather and illumination conditions.

A. SVM performance

The SVM classifier is tested, using a 9-fold-cross-validation, meaning each image is tested using the training information of the remaining 8 images. For proper classification and in order to avoid a bias towards one class label, the training sets has been balanced. However, the testing set will contain more shadow edges than reflectance edges, because they originate from the upstream shadow edge detection.

As depicted in Table I, the cumulated testing results yield an error rate of 19.25%. Note, that these point are already

TABLE I

9-FOLD-CROSS-VALIDATION. ROWS CONTAIN THE CORRECT CLASSES, COLUMNS THEIR CLASSIFICATION RESULTS.

	reflectance edge	shadow edge
reflectance edge	17093	3879
shadow edge	13700	56656

Total error rate: 19.25%
(correct: 73749, wrong: 17579)

the most ambiguous, having passed the shadow detection from Section II. For our purposes, more important than the overall error rate is, that the number of false positives could be reduced from 20972 (prior classification) to 3879 (using the proposed SVM classification). That means, 81.5% of all false detections could be rejected in the verification step.

B. Feature Selection

In section III, we proposed an image descriptor that is appropriate for the task of shadow edge detection. In order to benchmark its performance, the proposed feature set is evaluated in regard to its significance for the classification decision as well as its overall performance. For this purpose, we extracted features from different color channels and utilized multiple combinations of features in the RGB, HSV and CIE L*a*b* color spaces and computed their receiver operating characteristics (ROC) as depicted in Fig. 4.

Since our testing sets are not balanced, the achieved error rates of the best classification for different feature combinations according to Table I cannot be compared with each other. Hence, in order to compare the overall feature performance, we use the *equal error rate* and the *area under ROC* as shown in Fig. 4. The *equal error rate* yields the classification error for a balanced testing set and can be achieved as the intersection point of the ROC with a line of equal penalization for false classifications (gray lines in Fig. 4). The *area under ROC* is a popular measure to compare the overall classification performance since it incorporates all possible penalizations for a erroneous classification.

A feature evaluation from different color channels is shown in Fig. 4(a). In accordance with [12], the a^* channel performs significantly better than the b^* channel of the CIE L*a*b* color space. Our results also agree with [12], that features from color opponent perceptually uniform color spaces like CIE L*a*b* perform better than features from non opponent spaces like HSV. We realized an unexpected high performance of RGB features which is due to the fact that our application in scope are traffic scenes, where most of the shadow points lay on the gray road surface. For this special case, all color channels in the RGB color space have similar values. Any shadow edge will alter the local image statistic predominantly in the B channel. However, in arbitrary environments, the discriminance of the B channel is low and the CIE L*a*b* features are superior.

Fig. 4(b) shows the ROC curves and error measures for different feature combinations in the RGB, HSV and CIE L*a*b* color space. As mentioned before, RGB color

histograms significantly boost the feature performance for traffic scenarios. The proposed features (La+HSV+RGB hist) perform best in the whole ROC range and lead to a equal error rate of 19.0% (area under ROC: 0.88).

In a third experiment, we used local, global and point features from the CIE L*a*b* color space in combination with RGB color histograms, but omitted some of these feature types. As illustrated in Fig. 4c, a combination of different feature types leads to a more accurate classification with a decreasing equal error rate (only point features: 45%, point+local+global+histograms: 24.1%).

C. Shadow Edge Detection

Fig. 1 and Fig. 5 show the shadow detection results for multiple images containing challenging illumination, adverse road surface conditions and complex shadow geometries. The proposed approach yields a high precision with only few false positives. In particular, important reflectance edges like lane markings are treated correctly (e.g. Fig. 1 and Fig. 5(a),(c),(j)). Bitumen changes (Fig. 1), rails (Fig. 5(i)) and tar seams (Fig. 5(b),(g)) do not cause false detections. Note, that our approach only uses single color images and we do not need any camera calibration. The advantage of a non-parametric representation is that it can handle even complex shadow geometries (e.g. Fig. 5(c),(j)) where low-dimensional parametric descriptions like in [13] will fail.

However, the proposed detection algorithm still has difficulties with detecting some sharp, noticeable cast shadows (Fig. 5(i)), because the gradient magnitude response in the a^* channel is too high and the shadow edge candidate is not detected in the first stage. In addition, some wheel traces (Fig. 5(d)) could not be rejected by the SVM classifier. Whereas the former can be tackled with a different thresholding technique in the detection step, the latter can be improved by introducing more features in Section III.

VII. CONCLUSION AND FUTURE WORK

We proposed a novel approach for cast shadow recognition in single images from uncalibrated color cameras. The two-stage algorithm consists of a low-level detection of shadow edge candidates in the CIE L*a*b* color space, followed by a SVM classification as a verification step. The proposed features reduce false detections by 81.5% compared to the initial shadow edge detection. In addition, we introduce a non-parametric representation for complex shadow edge geometries which allows for the consideration of arbitrary shadow edge information in subsequent driver assistance systems. A broad evaluation on challenging traffic scenes yields good visual results with a high precision in detecting arbitrarily shaped cast shadows.

Future work will include an evaluation of color features that allow for improving the classification precision in the presence of wheel traces and other challenging disturbances.

VIII. ACKNOWLEDGMENT

The authors would also like to thank the Karlsruhe School of Optics and Photonics (KSOP).

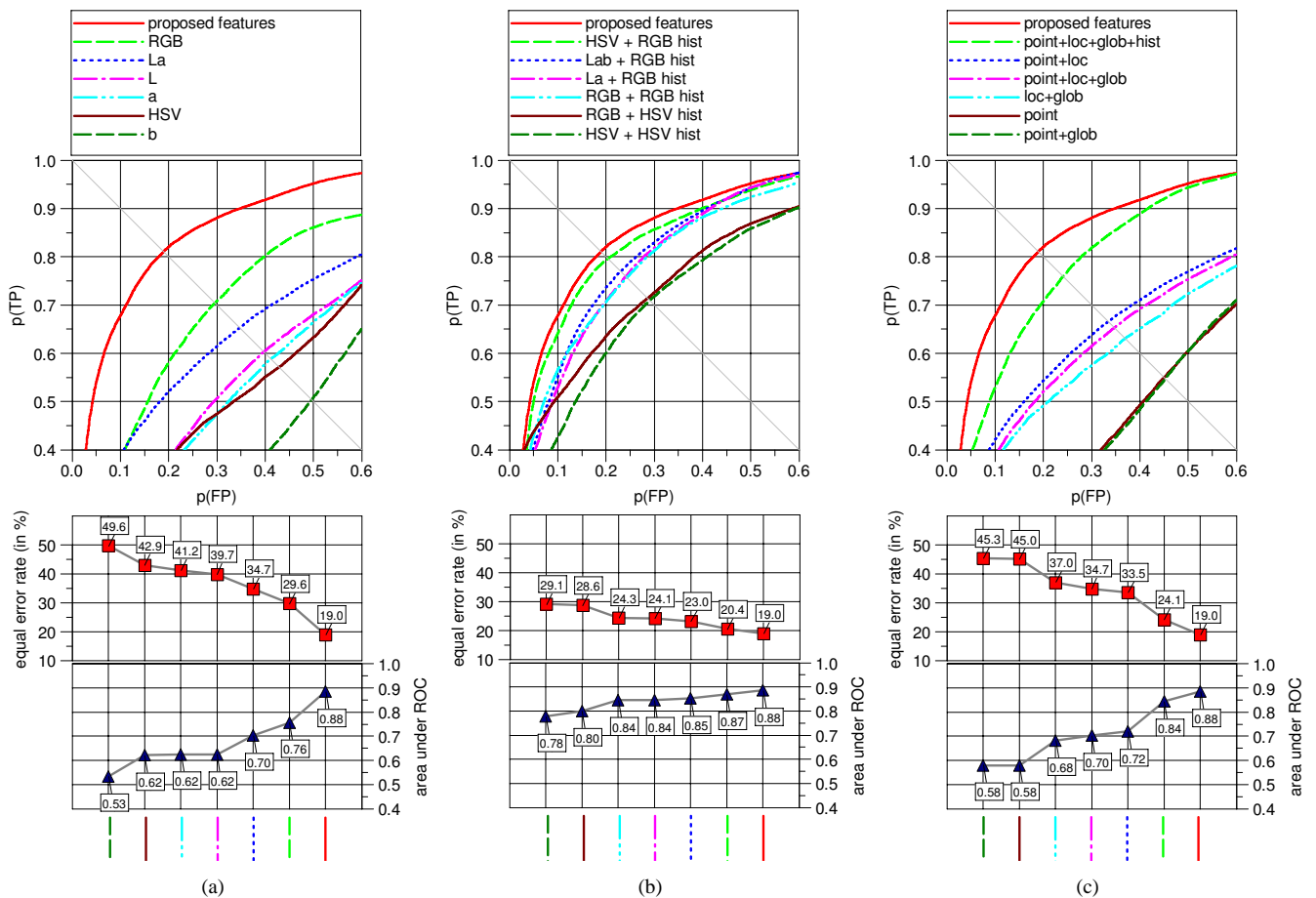


Fig. 4. **Feature selection.** (a) depicts the performance of selected color channels and spaces. (b) shows feature combinations extracted from the RGB, HSV and CIE $L^*a^*b^*$ color spaces and (c) illustrates the performance of CIE $L^*a^*b^*$ local, global and point features in combination with RGB color histograms when omitting single feature types. All experiments are compared to the proposed image descriptor (point, local and global features from L^*, a^*, H, S and V channel, as well as RGB color histograms). Note that only parts of the ROC diagrams are shown (top-left corner).

REFERENCES

- [1] Alexander Bachmann and Hildegard Kuehne. An iterative scheme for motion-based scene segmentation. In *IEEE Workshop on Dynamical Vision (in conjunction with ICCV '09)*, 2009.
- [2] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [3] Franklin C. Crow. Summed-area tables for texture mapping. *SIGGRAPH Comput. Graph.*, 18:207–212, January 1984.
- [4] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern classification*. Wiley, 2. ed. edition, 2001.
- [5] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *European Conference on Computer Vision (ECCV '02)*, 2002.
- [6] Graham Finlayson, Mark Drew, and Cheng Lu. Intrinsic images by entropy minimization. In *European Conference on Computer Vision (ECCV '04)*, 2004.
- [7] Graham Finlayson and Steven Hordley. Color constancy at a pixel. *Journal of the Optical Society of America: Optics, Image Science, and Vision*, 18(2):253–264, February 2001.
- [8] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian Conference on Computer Vision (ACCV '10)*, Queenstown, New Zealand, 2010.
- [9] Gevers, Stokman, T. Gevers, and H. M. G. Stokman. Classifying color transitions into shadow-geometry, illumination, highlight or material edges. In *IEEE International Conference on Image Processing (ICIP '00)*, pages 521–525. IEEE Computer Society Press, 2000.
- [10] Paul S. Heckbert. Filtering by repeated integration. *SIGGRAPH Comput. Graph.*, 20:315–321, August 1986.
- [11] M. Hussein, F. Porikli, and L. Davis. Kernel integral images: A framework for fast non-uniform filtering. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.
- [12] E.A. Khan and E. Reinhard. Evaluation of color spaces for edge classification in outdoor scenes. In *IEEE International Conference on Image Processing (ICIP '05)*, volume 3, pages III–952–5, 2005.
- [13] Jana Kosecká and Wei Zhang. Video compass. In *European Conference on Computer Vision (ECCV '02)*, pages 476–490, 2002.
- [14] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Estimating natural illumination from a single outdoor image. In *IEEE International Conference on Computer Vision (ICCV '09)*, 2009.
- [15] Jim Schwiagerling. *Field Guide to Visual and Ophthalmic Optics*, volume FG04. SPIE Publications, 2004.
- [16] A. Takahashi and Y. Ninomiya. Model-based lane recognition. In *IEEE Intelligent Vehicle Symposium (IV '96)*, pages 201–206, sep 1996.
- [17] Y. Weiss. Deriving intrinsic images from image sequences. In *IEEE International Conference on Computer Vision (ICCV '01)*, volume 2, pages 68–75 vol.2, 2001.
- [18] William M. Wells. Efficient synthesis of gaussian filters by cascaded uniform filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(2):234–239, 1986.
- [19] Hermann Winner, Stephan Hakuli, and Gabriele Wolf. *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*. Vieweg + Teubner, 1. auflage edition, Juli 2009.
- [20] Tai-Pang Wu and Chi-Keung Tang. A bayesian approach for shadow extraction from a single image. In *IEEE International Conference on Computer Vision (ICCV '05)*, volume 1, pages 480–487 Vol. 1, 2005.

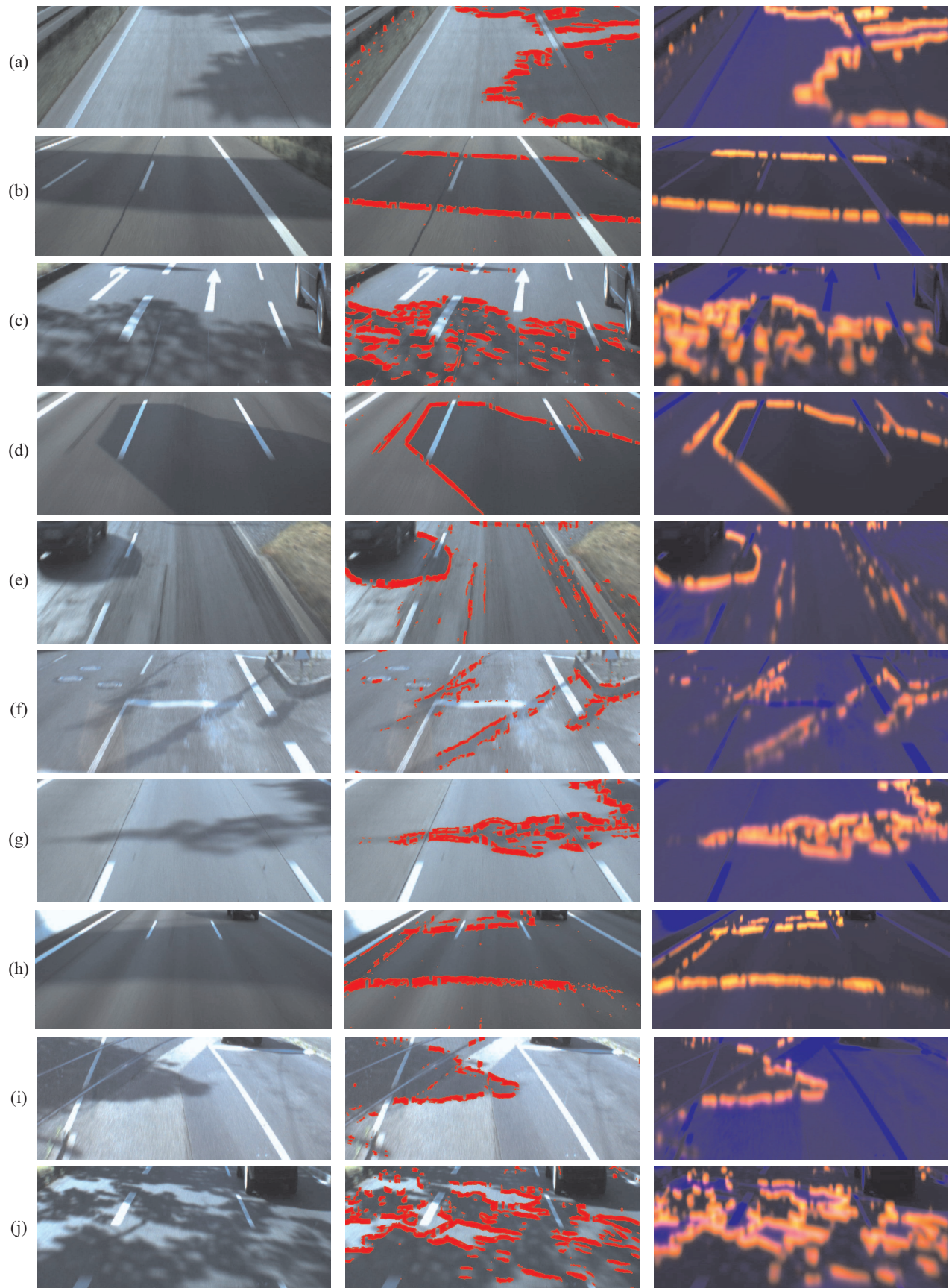


Fig. 5. **Shadow edge detection results.** Input images (left), verified shadow edge points (center) and non-parametric shadow edge density maps (right).