Real-time Physics-based Removal of Shadows and Shading from Road Surfaces

Bruce A. Maxwell Casey A. Smith Maan Qraitem Ross Messing Spencer Whitt
Nicolas Thien Richard M. Friedhoff
Tandent Computer Vision LLC
bmaxwell@tandent.com

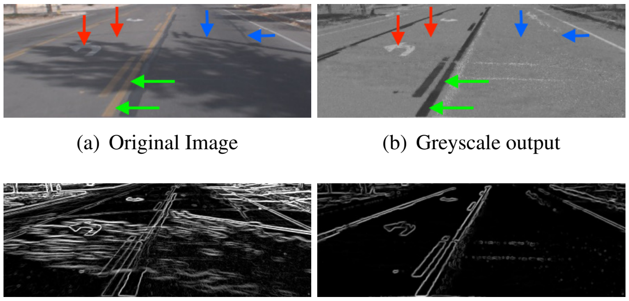
Abstract

We present a real-time physics-based system for generating an illumination free representation of road surfaces that maintains the distinction between asphalt and painted road markings. Cast shadows on road surfaces can create false features and modify the color of road markings, potentially masking important information for vehicle vision systems. We demonstrate a novel method for identifying the relative spectral properties of the direct and ambient illumination conditions and for using that to create an illumination-free 2D chromaticity space in log RGB. We then show how that representation can be used to generate an illumination-free greyscale representation that distinguishes road, white paint, and yellow paint, making it suitable for further analysis and classification. The entire process runs faster than 30Hz on 1 mega-pixel images using current automotive-grade embedded processing systems.

We evaluate the system on a paint detection task, comparing two types of learned classifiers, random forests and convolutional neural networks. For each type, one classifier is trained on the original images, and the other is trained on the illumination-free greyscale output. The classifiers are of identical complexity and trained on the same size data set. For both types, the classifier trained on the illumination-free outputs performs better, even on images with no cast shadows. The gap in performance is indicative of the cost of forcing a classifier to learn a task in the presence of the confounding illumination signal.

1. Introduction

Varied illumination and cast shadows on road surfaces creates a confounding signal that can mask or mimic road markings such as lane lines, turn arrows, crosswalks, and stop lines at intersections. These features are important for Advanced Driver Assistance Systems (ADAS) and autonomous vehicles; they convey information and can be useful for robust localization or free-space estimation.



(a) Original Image (b) Greyscale output

(c) Edges in original (d) Edges in greyscale output

Figure 1. (a) Original image. (b) Greyscale output where the road is grey, white paint is white, and yellow paint is black, irrespective of shadows and shading. (c) Edges in the original image. Illumination edges obscure the road marking edges. (d) Edges in the greyscale output. The road marking edges are clearly visible.

Illumination is a signal that can be arbitrarily complex, and it creates confusion at multiple scales and in multiple feature spaces, including intensity, color, and gradients. As shown in Figure 1a, white paint in shadow can be darker than lit asphalt (red arrows). Yellow paint in shadow, lit by blue skylight, can produce grey pixels with the same chromaticity as lit grey asphalt (green arrows). Shadow edges can be as sharp as paint edges and mimic the shapes of lane lines or other road markings (blue arrows). The confusion created by illumination is gone in the greyscale output of our system, as shown in Figure 1b.

One approach to a task such as road marking identification is to label a large data set and try to learn the task and illumination invariance simultaneously. However, the possible variation of cast shadows suggests that such a data set will be extremely large and costly to collect. While data sets such as KITTI [12] contain diverse illumination variation, training a network to learn robust illumination invariance for something like a road markings classification task requires many examples of illumination variation over all categories of road markings. In addition, any learned classifier must use part of its capacity to deal with illumination variation, a signal that is independent of the target task.

We present an alternative physics-based approach to identifying and separating illumination variation, treating illumination separately from the unrelated problem of classification. We describe a novel real-time system for generating an illumination-free representation of the road surface that discriminates between different reflectances, including asphalt and white paint. *Reflectance* is the intrinsic color of a surface, as if it were lit uniformly by white light. The algorithm requires only a small number of operations per pixel and runs > 30Hz on automotive computing hardware.

Figure 1 shows an example original image, an illumination free greyscale output, and the gradient magnitude outputs of the two images. The illumination edges in the original image mask the reflectance edges and the road markings are almost invisible. In the illumination-free edge image, the road marking edges are clear.

Our goal is not the exact recovery of reflectance, as would be used in true intrinsic image decomposition proposed in [2]. Rather, we seek to create an image representation which distinguishes the reflectances of interest (such as asphalt, white paint, and yellow paint) while removing or minimizing the influence of shadows and shading.

We evaluate our performance not based on constructed distance metrics comparing our outputs to some ground truth reflectance, but rather by measuring improvements on a downstream computer vision task. If the purpose of intrinsic imaging or shadow-free image representations is improving computer vision, the degree of that improvement is the most meaningful metric. In addition, any useful technique for dealing with shadows and shading within the ADAS domain must be faster than frame rate (i.e. 30Hz) on automotive computing hardware.

By separating illumination invariance from classification, the classification task becomes easier. The training set for identifying road markings needs to include variation only in geometry and the markings themselves, not variation in illumination. A classifier using our illumination free outputs should train faster, on less data, than a classifier trying to learn the same task on original images. Alternatively, a classifier of a given complexity trained and run on illumination free images should perform better than the same complexity classifier trained and run on the same number of original images. We demonstrate the latter case on a white paint detection task and compare several other state-of-the-art shadow removal techniques.

This paper has three novel contributions. (1) We present a novel automatic method for identifying the spectral relationship of the direct and ambient illumination in road scenes, (2) we present a real-time system for generating illumination-free outputs of road surfaces, and (3) we demonstrate that illumination-free outputs are better inputs for training a classifier on a real world task.

The methods and processes described herein are covered all or in part by the patents [21][18][26][27][28][29][30].

2. Background

This paper builds on prior work by [19]. They proposed a log chromaticity space that is invariant to illumination even for cases where the direct and ambient illumination are not the same color. This condition holds for most daylight situations where standard chromaticity shows significant variation between lit and shadowed areas. The challenge of using the log chromaticity of [19] is that it requires knowing the spectral ratio of the direct and ambient illuminants. Automatically detecting the spectral ratio is a difficult problem in general imagery. This work is the first to automatically detect the spectral ratio and to use log chromaticity space in a computer vision application.

Using illumination invariant or intrinsic images for computer vision tasks has long been proposed as a useful first step [2]. Early work on shadow removal includes the method of [9] who proposed a 1-D chromaticity space that is invariant to illumination changes under black-body illuminants. Extensions of the method can extend the results back to RGB [8] [10], but the loss of information in mapping the image into a 1-D space makes it difficult to differentiate reflectances that are brighter or darker versions of one another such as asphalt and white paint. For an example, see [7], which used the 1-D chromaticity space of [9] for a vehicle localization task. The 1-D mapping caused the white paint to disappear in road images. Despite that limitation, the work of [7] is one of the few cases that demonstrate the utility of illumination-free images over original imagery.

The 2-D log chromaticity space proposed by [19] provides an additional dimension of discriminability, suggesting it would provide even better performance on localization, but it requires an image-specific spectral ratio estimation, which has made it challenging to use.

Directly related work includes methods for identifying and removing shadows from images. A simple classifier to identify shadows on roads based on a small number of heuristic spatial and chromatic features is presented in [23]. They showed accuracies of 80-90%, but did not demonstrate the utility on a vision task. A learned decision tree classifier plus a conditional random field [CRF] to identify shadows on the ground in consumer photographs is used in [16]. While their approach is reasonable for certain types of shadows, it has difficulty with shadows that do not have well-defined boundaries or strong local gradients, such as soft shadows on roads. A segmentation approach to identify shadows is used in [13], but they specify their region growing method is also not robust to soft shadows.

A recent method for identifying and removing cast shadows is [24]. They trained a three-component convolutional neural network on cast-shadow/shadow-free pairs so that it learned to create a shadow-matte. Their goal was not to create illumination free images, as shading and fine scale illumination details remain in the output image, but to remove specific cast shadows. The network is also computationally demanding as a preprocessor for automotive tasks, requiring a high end GPU to achieve a 3Hz computation rate.

Intrinsic image generation is another potential approach to shadow removal. Examples of automatic intrinsic imaging systems include [11] [15] [1] [3] and [22]. Recently, [4] undertook a survey of numerous intrinsic imaging methods, and none of them were within two orders of magnitude of the speed necessary to be real time. The one exception, not tested by [4], was [22], which generates results at frame rate. However, their system makes strong assumptions about the scene, including that there is a single color illuminant so that standard chromaticity is useful. In an outdoor scene with primarily neutral surfaces of interest, that assumption is violated.

There is also work on estimating models of the sky in outdoor photos, which could be used to estimate spectral ratios. A model based on seeing part of the sky is estimated in [17]. More recently, [14] used a deep network to estimate sky model parameters without having to see the sky directly. While the latter approach holds promise, it is not sufficiently accurate for generating the log chromaticity representation of an image. Furthermore, nearby objects such as a painted building–not part of a sky model–can alter the ambient illumination.

3. Theory and Algorithms

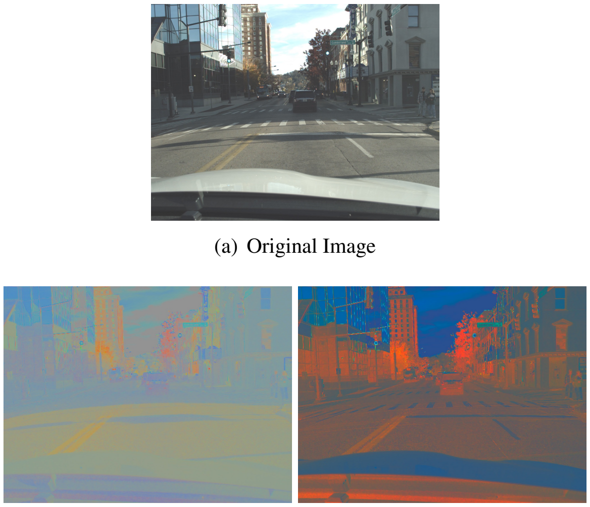
3.1. Log Space Chromaticity

The Bi-illuminant Dichromatic Reflection [BIDR] model is an extension of the Dichromatic Reflection model with an explicit ambient illuminant term [25][19]. The BIDR model makes two predictions about the colors of a uniform reflectance surface under natural illumination, such as occurs outdoors where the sun (yellow/red) is the direct illuminant and the sky (blue) is the ambient illuminant. First, the body reflection of a single color surface under varying illumination forms a line segment in linear RGB space offset from the origin. Second, the infinite extension of the line does not go through the origin. The latter condition implies that standard chromaticity (Equation 1) will not be illumination invariant. Figure 2 shows the difference between traditional chromaticity and the log chromaticity space for a typical urban scene.

$$(\hat{r}, \hat{g}) = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B} \right) \tag{1}$$

Following the derivation in [19], the appearance I of a single-color surface under varying illumination is

$$I = R_n(A + \gamma D) \tag{2}$$
$$\log I = \log R_n + \log(A + \gamma D) \tag{3}$$



(a) Original Image

(b) Standard Chromaticity

(c) Log Chromaticity

Figure 2. (a) Original image of an urban scene. (b) A standard chromaticity projection. The shadowed road surface is blue while the lit road surface is yellow and the white crosswalk is largely invisible. (c) The log chromaticity space. The road is a single color and the stop line and crosswalk are visible.

where A is the ambient illuminant, R_n is the reflectance of surface n, D is the direct illuminant, and γ is the proportion of the direct illuminant striking the surface. In linear RGB space (Equation 2), the appearance of each surface falls along a line with a unique length and orientation. In log space, the length and orientation of each reflectance's appearance curve, defined by the second term of Equation 3, is dependent only on the illuminants A and D. Different reflectance curves in log space, defined by R_n, are translated versions of one another. The ends of the curves are defined by $\gamma = 0$ (shadowed), and $\gamma = 1$ (fully lit).

The bright end minus the dark end of a reflectance curve in log space defines the direction of illumination variation over reflectances under that (A, D) illuminant pair. We define the normalized orientation of the endpoint difference (Equation 4) to be the *illumination spectral direction* [ISD].

$$\text{ISD} = \frac{\log I_{n,\text{lit}} - \log I_{n,\text{shadowed}}}{||\log I_{\text{n, lit}} - \log I_{\text{n, shadowed}}||} \tag{4}$$

While the mapping of the reflectance appearance lines from linear color space to log color space results in some curvature, the curves are very close to linear; almost all illumination variation is captured in one dimension. The remaining two dimensions constitute the log space chromaticity for that illumination condition.

Standard chromaticity is similar to log chromaticity with a neutral ISD (one where A and D have the same proportions of R, G, and B). Outdoors, A is dominated by the sky and is thus blue relative to D even on a mostly cloudy or hazy day. If we project along this non-neutral ISD to produce log chromaticity, the colored shadows are removed and reflectances which differ only in intensity (such as white paint and grey asphalt) are now distinct.

3.2. Identifying the Illumination Spectral Direction

The log space chromaticity transformation requires knowing the ISD, which depends on the relative spectra of A and D. Both A and D can change over time. In practice, both change slowly relative to frame rates. Therefore, it is not necessary to obtain an estimate for every image in a sequence. In addition, some images will not contain shadows, which must be handled properly.

The ISD is computable by identifying a shadowed pixel and a lit pixel corresponding to the same reflectance (e.g. asphalt). Our algorithm calculates features and uses thresholds based on physics, measurements of physical properties, and reasoning about the scene, which we assume is captured by a forward or rear-facing camera mounted on a car. The algorithm is applied to a trapezoidal region of interest [ROI] that corresponds approximately to the road surface. The steps of the algorithm are as follows.

```
1: Shrink the image to a width of no more than 150 pixels using 2x2 filters, keeping track of the percent variance of the pixels being averaged.
2: Generate a map of potential shadow pixels and dilate it.
3: Generate a map of potential lit pixels and dilate it.
4: Calculate an ISD at pixels likely to be on a shadow boundary with a valid entry in both the lit and shadow masks.
5: Filter out impossible or unlikely ISD estimates, including orientations that are close to neutral.
6: If there are too few ISDs, return zero confidence.
7: Use a robust estimator such as mean shift, to estimate an ISD from the remaining set of estimates.
8: Assign a confidence to the estimate based on the number of ISD estimates and the fraction that are inliers.
```

We incorporate the single frame estimates into a Kalman filter to reduce noise and maintain an ISD estimate when there are no shadows.

The algorithm contains a number of parameters, described below, including a definition of what constitute reasonable ISD values. Viewing the ISD as a unit vector, reasonable ISD values fall near an arc on the unit sphere defined by a neutral ISD $= (0.577, 0.577, 0.577)$ and a sunset ISD $= (0.789, 0.547, 0.299)$, which we measured close to sunset (very red sun) on a clear evening (deep blue sky). We exclude neutral orientations on that arc for two reasons: they would conflate asphalt and white paint, and a neutral orientation can only occur when the sky is fully cloudy and there would not be sharp shadows to interfere with identifying road markings. We define neutral orientations to be those with a dot product with neutral of greater than 0.9985. We define "near the arc" to be within a euclidean distance of 0.1 of the arc.

Our definitions of neutral and sunlight normals are camera independent so long as the camera is operating in linear mode where pixels values are proportional to the number of photons hitting the sensor (our sunset ISD was calculated from a DSLR image, for instance). If the camera response is linear, the ISD calculation is also independent of both exposure and white balance.

Because we do not need an ISD estimate every frame, we can choose the parameters conservatively to favor false negatives over false positives. The parameters provided are for exact reproducibility; sensitivity analysis performed on the parameters indicates that moderate changes do not significantly impact the time-smoothed output of the detected ISD. Varying conditions and the presence of other cars do not degrade the outputs (see supplementary materials).

- Potential shadow pixels are defined as having (a) a low percent variance (2%) in each color band and (b) a color that is roughly neutral or blue, but not more blue than a neutral surface would be at the measured sunset ISD.
- The shadow pixels are dilated by a trapezoid whose width is 8% of the ROI.
- Potential lit pixels are defined as having (a) a low percent variance (2%) in each color band and (b) no color band more than 45% brighter than another.
- The lit pixels are dilated by a trapezoid whose width is 4% of the ROI.
- Potential shadow boundary pixels must have a minimum log gradient magnitude of 0.2 and be a local maximum in the gradient image.
- Lit - shadowed must be at least 0.3 in log space in all channels.

This process is tailored to road images and is designed to be conservative: an ISD will likely only be detected when shadow boundaries of significant length cross a single-reflectance region of grey road. The previously detected ISD maintained by the Kalman filter is used on frames where no high-confidence shadow edge is detected. This algorithm will not work in a highly complex general scene.

The algorithm and runs at 500Hz on an NVidia Jetson TX1 for 1 megapixel images. Manual inspection of the results suggests that it works well on images of roads in a variety of conditions, whether mid-day or near sunset, clear sky or mostly cloudy/hazy. The evaluation results demonstrate that the ISD estimation is good enough to improve performance on a specific computer vision task. Other methods of identifying the ISD, including methods based on machine learning, may work as well or better. The real-time nature of the application, however, strongly constrains the complexity of potential ISD detection algorithms.

3.3. Greyscale Projection [GP]

Given an ISD, there are many possible illumination invariant outputs that could be generated for further processing. For instance, the raw 2D log chromaticity values can be used directly. For this evaluation and task, we use a greyscale output designed to highlight white and yellow paint on the road surface.

Given a non-neutral spectral ratio along the daylight axis, white paint projects to a more blue chromaticity than asphalt, and yellow projects to a much less blue chromaticity. Therefore, the blue direction on the log chromaticity plane is a useful axis to use for creating a greyscale output that separates asphalt, white paint, and yellow paint.

Given an ISD $\vec{N}$, the projection axis $\vec{N}^{\perp}$ which is perpendicular to $\vec{N}$ and aligned in the blue direction is given by Equation 5. N_b is the blue value of $\vec{N}$.

$$\vec{N}^{\perp} = (0, 0, 1) - N_b * \vec{N} \tag{5}$$

$\vec{N}^{\perp}$ needs to be computed only once per image or less often if $\vec{N}$ does not change between images. The dot product of each pixel value $\vec{P}$ in log space with $\vec{N}^{\perp}$ produces an illumination invariant greyscale value V_{raw}.

$$V_{\text{raw}} = log(\vec{P}) \cdot \vec{N}^{\perp} \tag{6}$$

Changes in $\vec{N}$ or the camera exposure will cause the output exposure and contrast to vary if we use V_{raw} directly. In order to produce consistent outputs, we pass the intermediate V_{raw} values through a piecewise linear transform to normalize the exposure and contrast of the greyscale output.

Let M be the median value of V_{raw} in the road trapezoid (fast approximations are sufficient): this provides the expected value of V_{raw} for the road surface. Let the contrast scale S be $(log(2), log(2), log(2)) \cdot \vec{N}^{\perp}$. This is the expected difference in V_{raw} between asphalt and white paint if white paint is twice as bright as asphalt.

The piecewise scaling to produce the greyscale projection V_{gp} is as follows:

$$V_{\text{gp}} = \begin{cases} 0.4 - ((M - S) - V_{\text{raw}}) * 0.075/S \\ \quad \text{if } V_{\text{raw}} \leq M - S \\ (V_{\text{raw}} - (M - S)) * .1/S + 0.4 \\ \quad \text{if } M - S < V_{\text{raw}} \leq M + S \\ (V_{\text{raw}} - (M + S)) * 0.075/S + 0.6 \\ \quad \text{if } V_{\text{raw}} > M + S \end{cases} \tag{7}$$

This is a 3-piece linear s-curve. Values much brighter than the median are given low slope to avoid clipping. Values between the median and the expected white paint color $(M + S)$ are given higher slope to improve contrast between asphalt and white paint. The transform is symmetric to provide similar contrast between asphalt and yellow paint on the dark end.

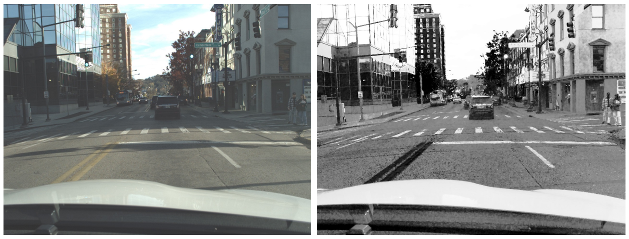

(a) Original Image (b) Greyscale Projection

Figure 3. (a) Original image and (b) an illumination invariant greyscale output where asphalt is grey, white paint is white, and yellow paint is black.

Transforming an image to the greyscale projection [GP] involves converting all RGB values to log rgb, computing a dot product, and scaling for contrast. This is three logs, four multiplications, and four additions per pixel. Each pixel is independent and can be processed in parallel to the degree possible on the target hardware. A straightforward implementation of the projection on an NVIDIA Jetson TX1 runs at over 150 fps, transforming every pixel in the original megapixel image. See Figure 3 for an example output.

4. Experiments and Results

We developed implementations on a CPU as well an NVidia Jetson TX1 GPU. The system runs faster than frame rate (¿ 30Hz) on megapixel-sized images on the Jetson, including all steps from Bayer interpolation through the final pixel transformation. It estimates the ISD on each frame and uses a Kalman filter to maintain an ISD over time.

To demonstrate the utility of illumination free images, we chose the task of identifying white paint on road surfaces. We labeled 304 images (244 training/60 test) of roads from car-mounted cameras. The data was sampled from 62 different video sequences taken with three camera sensors (Sony IMX224, Sony IMX035, Sony ICX285), four different lenses (from 60 to 180 degree field of view), in three different cities (Houston, TX; Knoxville, TN; Grand Junction, CO), at times of day including morning, noon, and near sunset, with conditions ranging from hazy and almost overcast to clear blue sky, in all four seasons with different foliage conditions.

We hand-labeled each pixel as road, white paint, or neither. We used 60 images sampled from video sequences not used in the training set as a test set. Figure 4 shows sample training images. A document with the test data and videos showing the system in action are in the supplemental.

We trained two types of classifiers: a random forest [RF] using Haar-like features and two configurations of convolutional neural networks [CNN]. Only white paint and road pixels were used for training and testing; pixels labeled as neither were excluded. We trained each classifier on origi-

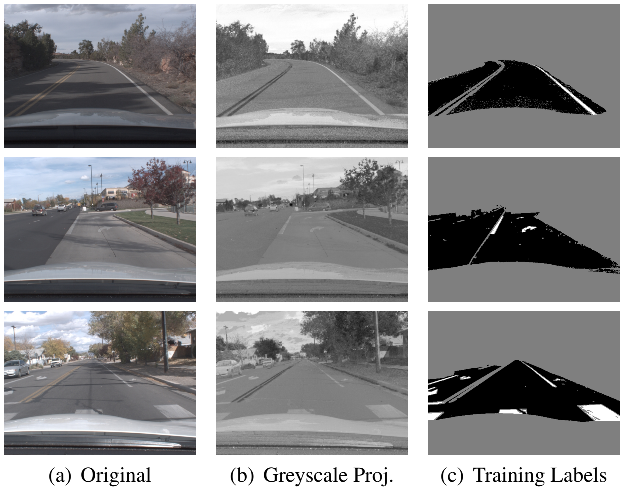


Figure 4. Training data used to train the classifiers. (a) Original image. (b) Greyscale projection. (c) Labels: road is black, white paint is white, all other pixels are grey.

nal and GP images. Each classifier used the same size training set, the same feature sets, and the same training parameters. Our goal was to test the hypothesis that by removing the confounding illumination signal, a classifier of the same complexity would improve its performance. The GP process is equivalent to one 2x3 filter, so the pre-processing cost is negligible compared to even a simple CNN.

The RF used 20 trees, each trained to a depth of 10 on a randomly selected 25% of each image using a 41x41 max window. We also trained two CNN models. One model has an input layer of 11x11x(3 or 1) with two layers of convolution/pooling each with 8 filters size 2x2 and stride 1 and a pooling layer of size 2x2. Each layer uses RELU. The final output is connected to a Dense layer with 64 nodes and RELU activation which is connected to two output nodes with softmax and categorical cross-entropy as a cost function. The second model has a 41x41 input layer and three layers of convolution/pooling, but is otherwise identical. The optimization algorithm is Adam with learning rate 0.001. The training set is 994583 11x11 samples (967891 41x41 samples), which included all of the pixels labeled as white paint and a randomly selected 5% of pixels labeled as asphalt; paint pixels make up about 28% of the resulting training set. We used 4.5M samples from images in video sequences not in the training set as the test set with the percentage of white pixels data: 2.07%. The CNN was created using Keras. [5].

We trained the CNNs on the GP images and color sRGB imagery. The color CNN has three channels and requires more computation than the combined GP process and 1-channel CNN given the same size CNN.

Figure 5 shows the comparative results for all of the classifiers. The improvement in recognizing white paint by using our preprocessed imagery is substantial. For the RF, at 95% recall, the precision increases from 38% to 99.0%. The CNN results show a more interesting story. The best CNN performer at 95% recall is the 11x11 CNN trained on GP images. The 41x41 CNN trained on GP images does slightly better than the 41x41 CNN trained on sRGB, and the 11x11 CNN trained on sRGB is much worse. The better performing 11x11 CNN trained on GP images is *more than 15x less complex* than the best CNN trained on sRGB.

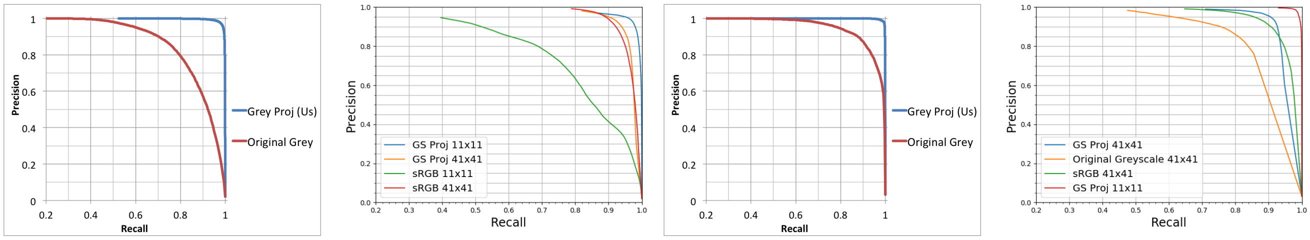
The improvement in recognition comes mostly from images with complex illumination variation, such as tree shadows. This is expected, as the GP should have no particular advantage in shadowless images. Figure 6 shows the RF performance on a test image with complicated shadows.

Interestingly, using the GP still improves performance on the seven shadow-free images in the test set, as shown in Figure 5 (c) and (d). Classifiers trained on original images have to cope with the confounding illumination signal, which must consume some capacity of the classifier or create similar inputs with different output categories. The illumination invariance the RF achieves is exceedingly poor (see Figure 6). However, by forcing the classifier to learn a degree of illumination invariance, it loses some of its classification ability even when there is no illumination variation.

As an additional test, we trained the RF using fewer training images. Figure 7 shows the precision-recall curves for the complete test set for the RF trained with the GP using data from just 4, 8 and 32 images, compared with the RF trained on sRGB using 32 and 244 (full data set) images. The RF trained on data from just 4 GP images roughly matches the performance of the RF trained on 32 or 244 sRGB images. The RF trained on just 8 GP images outperforms both of the sRGB RFs, and the RF trained on 32 GP images is almost identical to the RF trained on the full 244 GP images. Using the GP images means the classifier needs to learn only material intensity and geometry, not material intensity, geometry, and illumination.

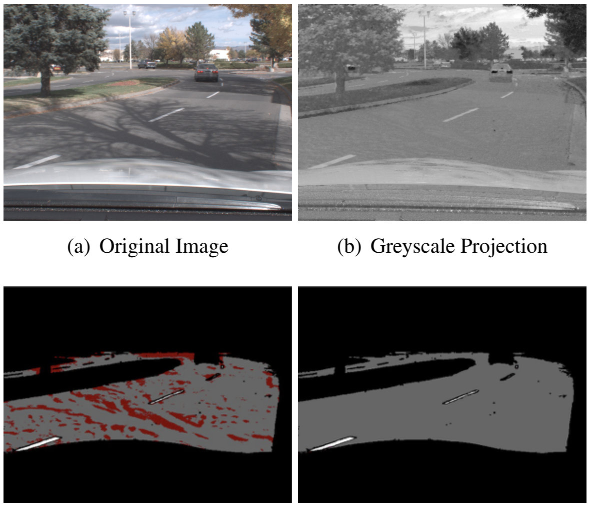
We also compare our results to existing state-of-the-art intrinsic imaging methods with published code, as well as standard chromaticity (normalized color) and the 1-D chromaticity in [9]. Other intrinsic image techniques require infeasibly long computational time and most often fail on outdoor imagery either by not removing shadows or by identifying white paint and road as having the same reflectance. Execution times on other methods range from 64s to 2.1 hours of cpu time per image. By comparison, a straightforward single-core implementation of the GP runs in 0.03s of cpu time per image. Additionally, both standard chromaticity and the chromaticity in [9], while fast to compute, fail to differentiate white paint and road. In contrast, the GP successfully removes shadows from outdoor scenes while still differentiating white paint and asphalt. See Figure 8.

The poor classification performance of the other meth-



(a) Random Forest (RF) (b) CNN (c) RF, test images lacking shadows (d) CNN, test imgs. lacking shadows

Figure 5. Precision/Recall curves for detecting white paint using classifiers trained on original images and GP outputs. (a) Random forest. (b) 11x11 CNN. (c) and (d) show curves for the RF and CNN classifiers detecting white paint on only test images with no shadows.



(a) Original Image (b) Greyscale Projection

(c) Classified Original (d) Classified Grey Proj

Figure 6. (a) The original image with a challenging shadow. (b) The greyscale projection. (c) The output of the RF classifier trained and run on original images. (d) The output of the RF classifier trained and run on the GP output. The classified images are shown at 90% recall. Grey: correctly classified road pixels. White: correctly classified white paint. Red: road misclassified as white paint. Magenta: white paint misclassified as road.

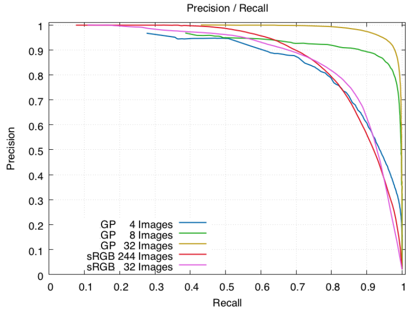


Figure 7. Precision-recall curves of the RF classifier trained on different numbers of training images and input types.

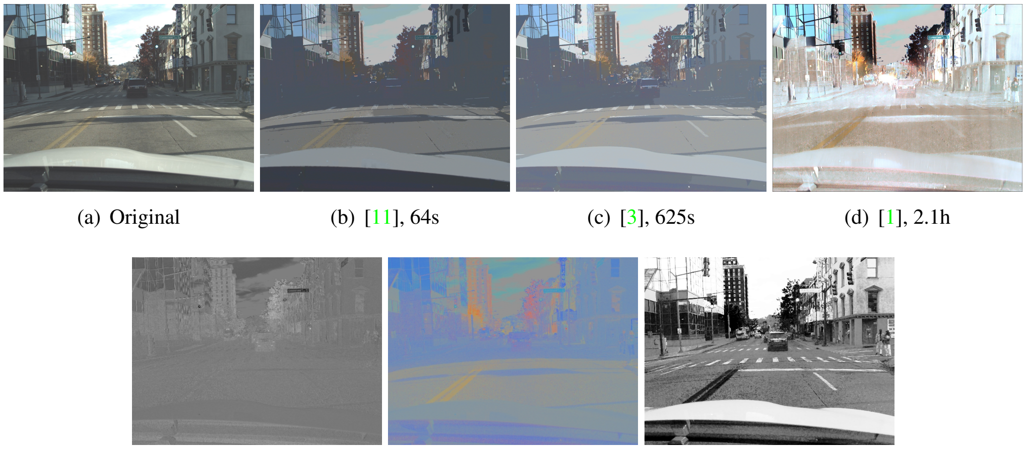
ods and of standard chromaticity is due either to their predictable failure to differentiate white paint and asphalt or their failure to remove shadows: they perform much worse than using the original images. Figure 9 shows the precision/recall plots of these methods compared to original sRGB images and the GP. All curves were generated by training CNNs with identical parameters, using the same source images, preprocessed by the technique indicated.

5. Discussion

Convolutional neural networks and advances in computational capability have greatly enhanced performance on many computer vision tasks. However, collecting and labeling data is expensive. Furthermore, recent work indicates that CNN performance is improving only logarithmically for linear increases in data size [31]. Intrinsic or illumination free images generated by a separate process provide an alternative path for improving performance. For vision tasks such as road marking identification, localization, or free-space estimation, the reflectance contains the necessary information without the confounding signal of illumination. Our physics-based computationally lightweight approach offers an alternative path to improving performance that is orthogonal to collecting more data.

Performing physics-based illumination analysis of automotive data is valuable but has been under-investigated partially because of a lack of good data. It requires physically meaningful images, free from sharpening, contrast enhancement, compression, tone-mapping, and other modifications applied by cameras intended for human image consumers instead of computer vision. All the major modern automotive cameras are capable of being configured to capture data appropriate for physics-based analysis. There are no technical barriers to obtaining physically accurate data in a production automotive environment, and all of the data in this study was captured using off-the-shelf hardware.

The standard publicly available data sets for automotive computer vision are not appropriate for physics-based vision in one or more ways. Two of the best known data sets are KITTI [12] and CityScapes [6]. KITTI avoids many non-linear color distortions, but the autoexposure used by its cameras clips many pixels, permanently destroying ac-



(a) Original (b) [11], 64s (c) [3], 625s (d) [1], 2.1h

(e) [9], 0.06s cpu (f) Std. Chrom., 0.01s cpu (g) Our Projection, 0.03s cpu

Figure 8. Top row: Sample results and processing times for intrinsic image methods with published code. Bottom row: 1-D log of chromaticity projection, standard chromaticity, and our GP. The intrinsic imaging methods require infeasible computation times, and the other chromaticity methods fail to distinguish asphalt and white paint.

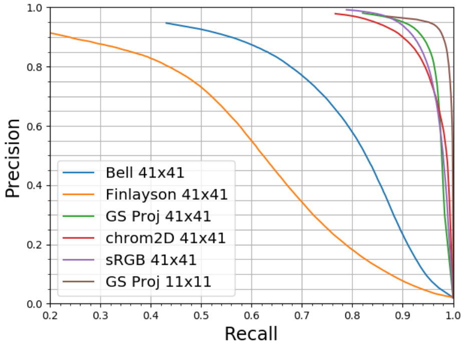
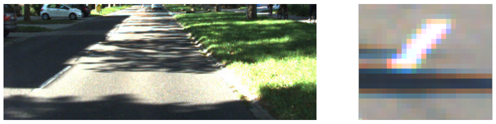


Figure 9. Precision/Recall curves for the CNN classifiers detecting white paint comparing different preprocessing techniques. "Bell" is [3] and "Finlayson" is our implementation of [9].



(a) Overexposed and clipped (b) False Colors

Figure 10. Examples from the KITTI data set. (a) Overexposed and clipped pixels. (b) Poor-quality de-Bayering that generates vibrant false colors along boundaries.

curate color data, and the de-Bayering algorithm leaves artificial neon colors on all sharp boundaries, including shadows, as shown in Figure 10. CityScapes uses high quality de-Bayering, but the 16-bit HDR data is not actually linear. Physically correct data should provide very similar ratios of the ambient and direct illumination when shadow boundaries cross multiple nearby reflectances. However, in CityScapes, measuring the ISD on white paint gives a different value than measuring the ISD on nearby asphalt.

KITTI and Cityscapes are excellent data sets with many uses, but they are unsuitable for physics-based algorithms.

6. Summary

We present a real-time system for generating illumination invariant imagery for automotive applications. We describe a means of automatically characterizing the illumination conditions with the *illumination spectral direction* [ISD] which defines the chromatic relationship between direct and ambient light sources. Projecting out this direction in log(RGB) space produces an illumination-invariant chromaticity space. We show how to produce a greyscale projection of the original image which is free of shadows and shading but still differentiates asphalt and white paint.

We evaluated the utility of illumination-free images on the task of distinguishing white paint from road. We trained two types of classifiers, comparing versions trained on original images with versions trained on the illumination free greyscale projection. Detecting white paint was easier once shadows and shading were removed, as demonstrated by a substantial improvement in recognition performance. The greyscale projection also performed better than state-of-the-art intrinsic imaging techniques, all of which are additionally too slow to use in an automotive environment. The classifier trained on the greyscale projections performed better *even on images with no shadows*. In addition, the classifier using the greyscale projection worked better even when trained on substantially less training data. We conclude that compact physical modeling prior to machine learning can be beneficial.

References

[1] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Trans. on Pattern Analysis and Mach. Intelligence*, 37(8):1670–1687, Aug 2015. 3, 8

[2] H. G. Barrow and J. M. Tenenbaum. *Computer Vision Systems*, chapter Recovering intrinsic scene characteristics from images, pages 2–26. Academic Press, 1978. 2

[3] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Trans. Graphics (SIGGRAPH)*, 33(4), 2014. 3, 8

[4] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. Intrinsic decompositions for image editing. *Computer Graphics Forum*, 36(2):593–609, 2017. 3

[5] François Chollet et al. Keras. `https://keras.io`, 2015. 6

[6] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 7

[7] Peter Corke, Rohan Paul, Winston Churchill, and Paul Newman. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2085–2092. IEEE, 2013. 2

[8] G. D. Finlayson, M. S. Drew, and L. Cheng. Intrinsic images by entropy minimization. In T. Pajdla and J. Matas, editors, *Proc. of European Conf. on Computer Vision*, LNCS 3023, pages 582–595, 2004. 2

[9] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *Proc. of European Conf. on Computer Vision*, pages 823–836, London, UK, 2002. Springer-Verlag. 2, 6, 8

[10] Graham D. Finlayson, Steven D. Hordley, Cheng Lu, and Mark S. Drew. On the removal of shadows from images. *IEEE Trans. on Pattern Analysis and Machine Vision*, 28(1):59–68, 2006. 2

[11] Elena Garces, Adolfo Munoz, Jorge Lopez-Moreno, and Diego Gutierrez. Intrinsic images by clustering. *Computer Graphics Forum*, 31(4):1415–1424, 2012. 3, 8

[12] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. 1, 7

[13] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2956–2967, Dec 2013. 2

[14] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2017. 3

[15] Vivek Kwatra, Mei Han, and Shengyang Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In *International Conference on Computational Photography*, 2012. 3

[16] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *European Conference on Computer Vision*, 2010. 2

[17] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98(2):123–145, 2011. 3

[18] Bruce A Maxwell and Richard M Friedhoff. Bi-illuminant dichromatic reflection model for image manipulation, 2015. US Patent 8,976,173. 2

[19] Bruce A Maxwell, Richard M Friedhoff, and Casey A Smith. A bi-illuminant dichromatic reflection model for understanding images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 2, 3

[20] Bruce A Maxwell, Richard M Friedhoff, and Casey A Smith. Bi-illuminant dichromatic reflection model for image manipulation, 2015. US Patent 8,976,174.

[21] Bruce A Maxwell, Casey A Smith, and Richard M Friedhoff. Method and system for separating illumination and reflectance using a log color space, 2009. US Patent 7,596,266. 2

[22] Abhimitra Meka, Michael Zollhöfer, Christian Richardt, and Christian Theobalt. Live intrinsic video. *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, 35(4), 2016. 3

[23] S. Park and S. Lim. Fast shadow detection for urban autonomous driving applications. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1717–1722, Oct 2009. 2

[24] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[25] Steven A. Shafer. Using color to separate reflection components. *Color Research Applications*, 10:210–218, 1985. 3

[26] Casey A Smith. Method and system for classifying painted road markings in an automotive driver-vehicle-assistance device, 2015. US Patent 9,218,534. 2

[27] Casey A Smith. Method and system for classifying painted road markings in an automotive driver-vehicle-assistance device, 2016. US Patent 9,361,527. 2

[28] Casey A Smith. Method and system for classifying painted road markings in an automotive driver-vehicle-assistance device, 2016. US Patent 9,466,005. 2

[29] Casey A Smith. Method and system for classifying painted road markings in an automotive driver-vehicle-assistance device, 2018. US Patent 9,875,415. 2

[30] Casey A Smith. Method and system for classifying painted road markings in an automotive driver-vehicle-assistance device, 2018. US Patent 10,032,088. 2

[31] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. 7