ADAPTIVE SHADOW REMOVAL FOR VARIABLE ENVIRONMENTAL CONDITIONS

A Dissertation
Presented to
The Academic Faculty

By

Jay Danner

In Partial Fulfillment
of the Requirements for the Degree
Master of Electrical and Computer Engineering

Georgia Institute of Technology

May 2017

ADAPTIVE SHADOW REMOVAL FOR VARIABLE ENVIRONMENTAL CONDITIONS

Approved by:

Dr. Linda Wills, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Lee Lerner, Advisor
Georgia Institute of Technology

Dr. Ayanna Howard
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Patricio Vela
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. James Hamblen
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: April 28, 2017

ACKNOWLEDGEMENTS

This is acknowledging things.

Table of Contents

List of Tables

List of Figures

SUMMARY

The motivation for this research is to provide a quantitative model for improving shadow detection in arbitrary environments. Many computer vision applications extract foreground pixels to capture moving objects in a scene. However, since shadows share movement patterns with foreground objects (and have a similar magnitude of intensity change compared with the background model), they tend to be extracted along with the desired object pixels. Shadows generally contribute to inaccurate object classifications and impede proper tracking of foreground objects. Contemporary shadow removal algorithms have made great strides in discriminating between object pixels and shadow pixels, but are hampered by a lack of scene-independence. In order to perform optimally, these leading methods require assumptions to be made about key factors of a scene, including illumination constancy, color content, and shadow intensity. As a result, no leading shadow removal method is robust enough to compensate for environmental change over time, nor are they suitable for deployment into a particular environment without a priori tuning of parameters.

This research evaluates popular shadow removal methods, extracts corresponding algorithmic parameters that affect shadow removal, correlates these parameters with salient environmental aspects, and finally leverages this correlation to improve shadow removal efficacy across diverse environments. Data collection and validation were performed using a collection of

widely-used computer vision datasets. Parameters, both algorithmic and environmental in nature, are identified, correlated, and evaluated using analytic tools. Particularly, a strong correlation is found using the average attenuation of dark foreground pixels in a frame. Exploiting this correlation, an adaptive model improves shadow detection by up to 10% and improves shadow-object discrimination by up to 28%. Additional indirect environmental factors are found to increase the effectiveness of this attenuation model. Brightness calculation methods are shown to improve attenuation correlation by 7% to 20%. Identifying low-contrast feature keypoints in a scene is also found to improve attenuation-correlation in some environments by up to 12%.

Chapter 1

Introduction

The need for accurate video analytics has emerged in many application areas, including surveillance, transportation management, and smart-city sensing infrastructure. Many of these applications, e.g., pedestrian and traffic counting [??], perform classification and analysis on moving objects. In order to detect and extract moving objects in an environment, foreground pixels are differentiated from those of the background through the use of statistical techniques, including Mixture of Gaussians, and Multi-Modal Mean [??, ??]. These strategies establish a model of the background of a scene, which changes gradually over time. This background model is then compared directly to a frame. By finding the difference between an image frame and its background model, a mask containing the foreground pixels is created. Foreground pixels are then grouped together and segmented as a moving object, which is analyzed according to the needs of the application.

Applications that analyze moving objects rely on the accuracy of the foreground extraction of foreground pixels. These applications are disadvantaged by the fact that shadows are often mischaracterized as foreground objects, and are included as part of a moving object. This is often because shadows possess similar movement patterns and brightness compared to non-shadow foreground objects [??]. Figure 1.1 illustrates the effect shadows have on the segmentation of foreground objects.

The inclusion of shadows in foreground objects may hamper detection and tracking in several ways. Prominent issues, noted by Sanin et al. [??], include the distortion of an object's appearance model (required to properly track an object), and the erroneous joining of multiple objects into one labeled connected-component. Additional details regarding a shadow's effect on tracking can be found in [??]. The removal of shadows from foreground objects is thus a vital step in accurately segmenting moving foreground objects.

Classifying shadow pixels within moving foreground objects has been approached in numerous ways, including color-based attenuation models, geometric projective models, and texture-matching models [??, ??, ??]. Machine-learning has also been employed to attempt to learn the appearance of cast shadows in an unsupervised manner [??, ??, ??]. A taxonomy of shadow removal methods produced by Prati et al. summarizes and evaluates four contemporary method classes: Statistical Nonparametric, Statistical Parametric, Deterministic Nonmodel-based and Deterministic Nonmodel-based
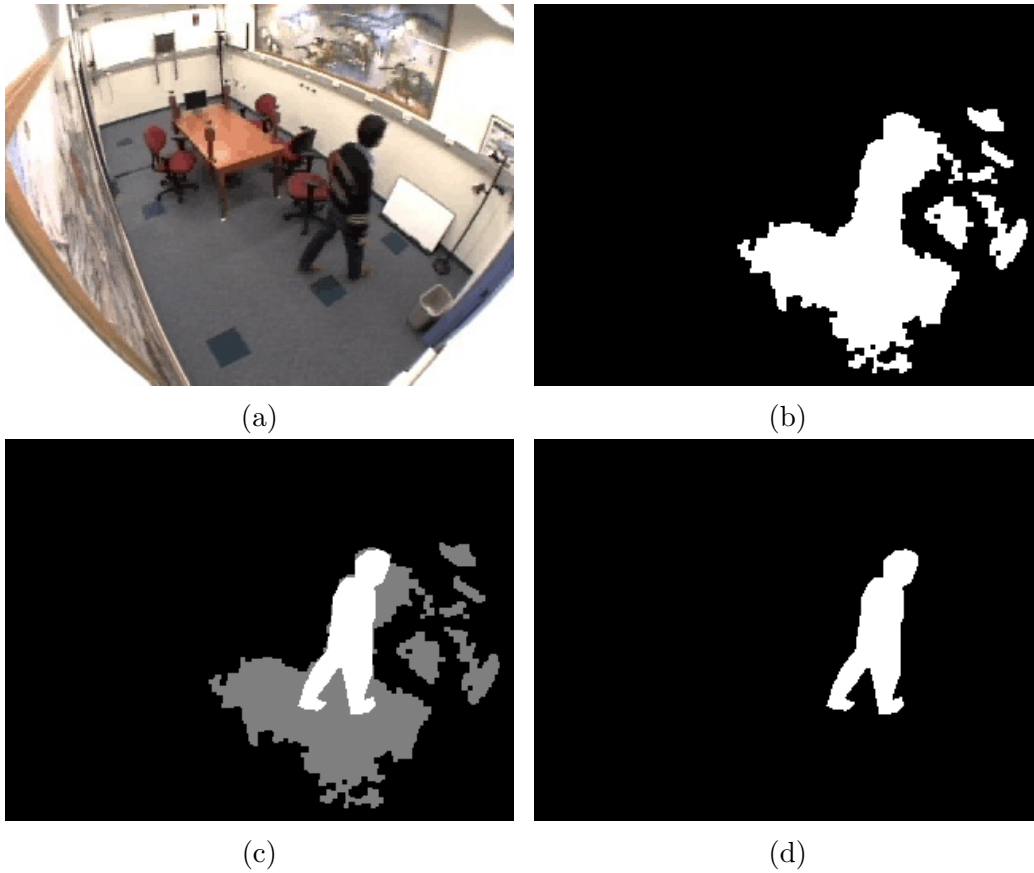
Figure 1.1: The need for shadow removal: (a) Original frame from the aton_room dataset [**??**]. (b) Result of foreground extraction, with shadows incorrectly being detected as moving objects rather than as background. (c) Research goal: proper segmentation into shadow regions and true foreground. (d) Proper identification of foreground with the removal of shadows.

[2]. The study concluded that the simpler methods were more suited for general practice, but "to detect shadows efficiently in one specific environment, [adding] more assumptions yields better results." A second algorithm survey conducted by Sanin et al. in [1] evaluated more modern methods (catalogued as Chromacity, Geometry, Physical, Small Region Texture, and their own contribution, Large Region Texture) on the same datasets as above, yielding similar results concerning the generalization of shadow removal to an arbitrary scene. Mitra et al. provides a survey of threshold selection strategies for identifying shadows in moving foreground objects [3].

1.1   The Problem

These surveys indicate that existing shadow removal algorithms fail to optimally adapt to varying environmental properties; these methods quantifiably benefit from assumptions made about key factors of a scene, including illumination constancy, color content, and shadow intensity. In order to facilitate optimization, shadow removal methods possess algorithm parameters that are manually tuned to an environment. The reliance on environmental assumptions affects shadow removal in two ways: firstly, shadow removal performs suboptimally when deployed in an arbitrary environment, and secondly, even when manually calibrated, shadow removal fails to adapt to environmental parameters that change over time. From an application context, a surveillance system that monitors a sun-lit environment throughout an entire

4

day will possess a wide range of shadow qualities when comparing shadows cast at dawn to shadows cast in the evening. Shadows cast in the same location may vary in darkness, size, orientation, and shape depending on the time of day. Therefore shadow removal must adapt not only to diverse environments, but continually adapt as environmental properties vary over time.

We quantitatively demonstrate several cases that indicate the need for adaptation. Shadow removal is judged by its detection rate and its discrimination rate, further detailed in section 2.1. Detection rate indicates the number of shadow pixels correctly identified, while discrimination rate indicates the number of foreground object pixels that are correctly identified. Figure **??** modifies a parameter belonging to Physical shadow removal, which controls the chromacity range in which a shadow can lie. The parameter (*coneR1*) causes Physical shadow removal to perform optimally in the aton_highway1 frame at a value of 0.2. Similarly, the same algorithm performs poorly in the included dataset aton_campus with the same parameter value. However, if *coneR1* is modified from 0.2 to 0.68, aton_highway1 experiences a 77% loss of detection, while aton_campus gains 53% discrimination in exchange for a 15% loss of detection (1.3). Geometry shadow removal was found to also showcase different results on the same scene, but with differing parameters (Figure 1.4).
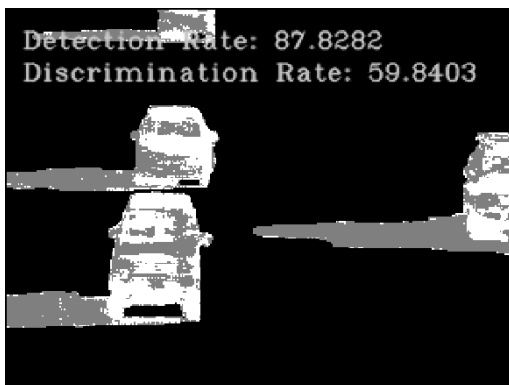
(a) aton_highway1
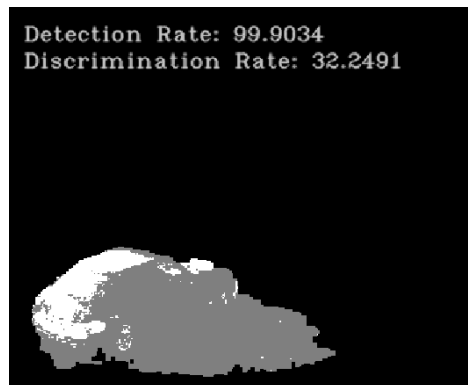
(b) aton_campus

Figure 1.2: *coneR1* set to 0.2. (a) Detection: 87.8, Discrimination: 59.84. (b) Detection: 99.9, Discrimination: 32.2
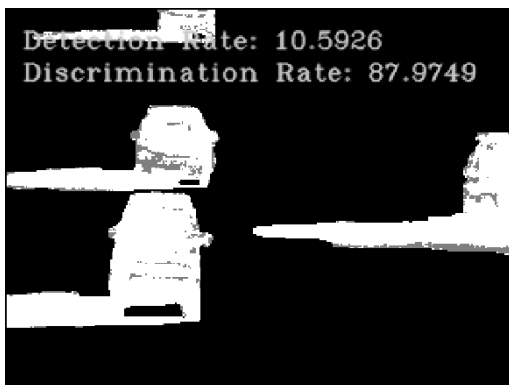
(a) aton_highway1
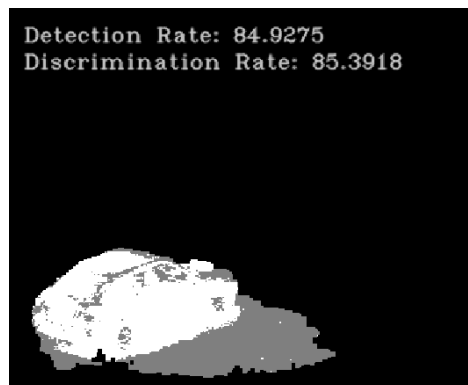
(b) aton_campus

Figure 1.3: *coneR1* shifted from '0.2' to '0.68.' (a) Detection: 10.59, Discrimination: 87.97. (b) Detection: 84.9, Discrimination: 85.4. No parameter value provides consistent accuracy across datasets.

(a) *gWeight: 70*   (b) *gWeight: 0*
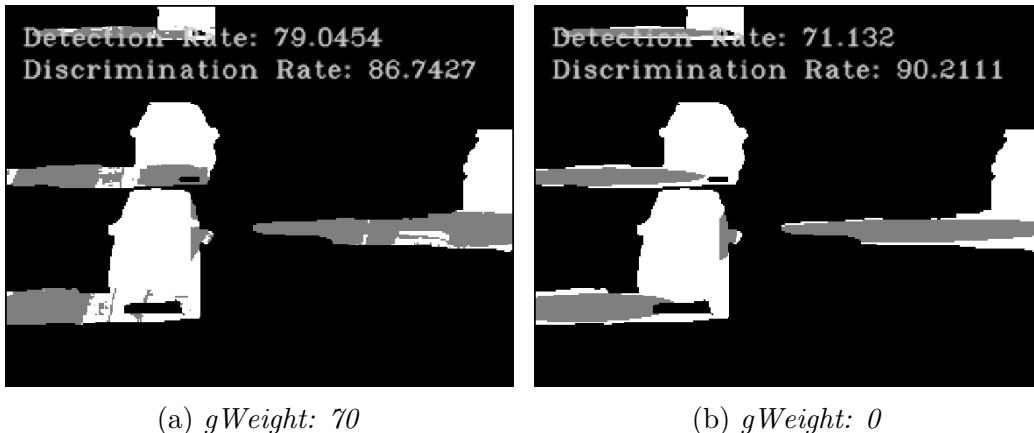
Figure 1.4: *gWeight* shifted from '70' to '0.' (a) Detection: 79.05, Discrimination: 86.74. (b) Detection: 71.13, Discrimination: 90.21

1.2   Objective and Contributions

Our research seeks to establish an understanding of environmental properties that affect shadow removal, and use that understanding to optimize shadow removal in an arbitrary environment. This is achieved by automatically calibrating an algorithm's parameters based on observed environmental properties. Furthermore, we seek to create an understanding of how these environmental properties change over time, in order to continuously adapt shadow removal algorithms. These objectives require the creation of an adaptive model that automatically configures a shadow removal method to optimally perform given the observed environmental properties.

Our research makes the following contributions. We perform a qualitative assessment of each algorithm's performance in various environments. We construct and utilize an interactive framework for evaluating the sensi-

7

tivity of an algorithm with respect to its mutable parameters. We identify and quantify observed environmental properties, and correlate them to sensitive algorithm parameters. Finally, we demonstrate the construction of an adaptive model, capable of leveraging correlated environmental properties to automatically tune an algorithm's parameters. The demonstration is completed by constructing a proof-of-concept model for Physical shadow removal.

Our proof-of-concept adaptive model draws upon the correlation between brightness attenuation in shadow regions and one of the Physical algorithm's parameters, *coneR1*. It improves shadow detection by up to 10% and shadow discrimination by up to 28% on a range of standard datasets from ATON and PETS [**??**, **??**]. Additional indirect environmental factors are found to modify the effectiveness of the adaptive model. Various brightness calculation methods are shown to influence attenuation correlation by 7% to 20%. A study of low-contrast feature keypoints in a scene was shown to improve attenuation-correlation by up to 12% in some environments.

The outline of this thesis is as follows. In Chapter 2, we detail the shadow removal algorithms utilized in this study, and outline the steps taken to produce an adaptive model. Chapter 3 assesses the shadow removal algorithms for their performance in diverse environments, as well as their sensitivity to parametric change. These assessments culminate in the construction of a proof-of-concept adaptive model for Physical shadow removal, using its *coneR1* parameter. Chapter 3 also explores indirect environmental proper-

ties and their potential impacts on the performance of the adaptive model. Chapter 4 quantifies and discusses the results of the adaptive model, and the indirect environmental properties' effects on the model. Chapter 5 summarizes conclusions and future work.

Chapter 2

Approach

This chapter details evaluation metrics, provides background on the surveyed shadow removal algorithms, and outlines our research methodology. In section 2.1, we quantify the metrics used to evaluate the accuracy of shadow removal. Section 2.2 provides background on the shadow removal algorithms examined in this study. Finally, our research approach is outlined in section 2.3.

2.1   Evaluation Metrics

The efficacy of a shadow removal algorithm on a given frame is evaluated with the popular metrics *Detection* (η) and *Discrimination* (ξ) (Eqn. 2.1, Eqn. 2.2) [??]. These measure how many shadow pixels are correctly identified, and how many foreground object pixels are correctly preserved, respectively.

The metrics are calculated using true positives (TP) and false negatives (FN) of both foreground pixels and shadow pixels. Subscripts signify identifications on shadow pixels (S) or foreground pixels (F).

$$\eta = \frac{TP_S}{TP_S + FN_S} \tag{2.1}$$

$$\xi = \frac{TP_F}{TP_F + FN_F} \tag{2.2}$$

2.2 Background: Shadow Removal Methods

We select five popular shadow removal algorithms for this study, named Chromacity, Physical, Geometry, Small-Region Texture, and Large-Region Texture. These shadow removal methods use differing techniques for identifying foreground shadow pixels. These techniques are color information (Chromacity), brightness attenuation (Physical), shadow shape (Geometry), and texture information (Small-Region Texture, Large-Region Texture). We used standardized implementations of the shadow removal methods, licensed under GPL v3+ and written in C++ [**??**].

2.2.1 Chromacity

Chromacity, or *hue*, represents the base color of a pixel, and is separable from brightness and saturation. Chromacity-based shadow removal methods

maintain that a pixel, when covered by a shadow, loses luminosity (or brightness), while retaining its chromacity. This assumption is referred to as color constancy, or linear attenuation. This study implements one such algorithm from Cucchiara et al. [**??**], implemented using the Hue-Saturation-Value (HSV) color representation.

Cucchiara et al. observe a shadowed pixel in the foreground (fg_p) retains its hue when compared to the corresponding background pixel (bg_p), while losing saturation and intensity (value). In order to be considered a shadow, the hue, saturation, and value of fg_p must fall within the pre-determined thresholds τ_H, τ_S, and $[\beta_1, \beta_2]$ (Eqn. 2.3, 2.4, 2.5).

$$|fg_p(H) - bg_p(H)| \leq \tau_H \tag{2.3}$$

$$fg_p(S) - bg_p(S) \leq \tau_S \tag{2.4}$$

$$\beta_1 \leq \frac{fg_p(V)}{bg_p(V)} \leq \beta_2 \tag{2.5}$$

The thresholds are optimized for the environment of the algorithm's intended application. Due to its reliance on curated thresholds, Chromacity shadow removal is sensitive to strong illumination changes. Furthermore, the assumed linear attenuation model performs worse with dark shadows.

2.2.2 Physical

For a shadow pixel to attenuate linearly, the light source casting the shadow must consist of primarily white light. Many environments experience multiple light sources, whether they are the sun, surface reflections, or blue light refracted from the sky. The presence of non-white light sources causes non-linear attenuation from a foreground shadow pixel to its background pixel.

This study uses an implementation of Cheng et al.'s Physical shadow removal, which attempts to learn the attenuation model of a shadow using a Gaussian Mixture Model (GMM) [**??**]. Three features are used to learn the attenuation model of a pixel (p): illumination attenuation (α_p), red-green direction (θ_p), and blue direction (ϕ_p). θ_p and ϕ_p are spherical coordinates, derived from the representation of the pixel p as a vector $\vec{v}_p$ in the RGB coordinate plane. $\vec{bg}_p$ represents the pixel vector associated with the corresponding background model. Eqn. 2.6, Eqn. 2.7, and Eqn. 2.8 describe the calculation of these features.

$$\alpha_p = \frac{||\vec{v}_p||}{||\vec{bg}_p||} \tag{2.6}$$

$$\theta_p = arctan(\frac{\vec{v}_p(G)}{\vec{v}_p(R)}) \tag{2.7}$$

$$\phi_p = arccos(\frac{\vec{v}_p(B)}{||\vec{v}_p||}) \tag{2.8}$$

13

Physical shadow removal first uses a weak detector to identify candidate shadow pixels, calculates the appropriate color features, and uses them to update the GMM. The GMM learns the attenuation model of a shadow over time, and is used to discriminate between foreground object pixels and shadow pixels.

2.2.3   Geometry

Widely-used Geometry-based shadow removal methods attempt to identify shadow regions in a foreground object using projective geometry [**??**]. The implementation evaluated in this study, proposed by Hsieh et al. [**??**], characterizes the geometric moments of foreground blobs in an attempt to identify the vertical peak and center of gravity of the objects. Using this information, the foreground object is split into an object region and a shadow region.

Geometric removal methods often require that a shadow and an object retain a clear orientation in regard to one another. Geometric removal is best deployed in environments with distinct, upright objects with a strong directional source of light.

2.2.4   Small-Region Texture

Texture-based shadow removal attempts to match shadow pixels based on the underlying background texture, i.e., structural patterns observed in both the background model and the foreground. If similar structural patterns are

observed, it is concluded that the foreground region does not occlude the background, and is therefore more likely to be a cast shadow.

Small-Region Texture (SRT) shadow removal, proposed by Leone et al. [**??**], utilizes a set of Gabor functions with various bandwidths, orientations, and phases. A set of candidate shadow pixels, determined by a weak detector similar to that of Physical removal, is projected onto the set of Gabor filters. After analyzing both foreground and background, the texture correlation is found by computing the Euclidean distance.

2.2.5 Large-Region Texture

Large-Region Texture (LRT) shadow removal recognizes that the small regions analyzed using SRT may fail to contain enough structural information to match foreground to background. LRT takes advantage of Chromacity removal to produce regions of probable shadow candidates, and correlates the gradient information of both the foreground and background regions. LRT removal proves most effective in environments characterized by strong textural features and large contiguous shadow regions.

2.3 Approach

Our approach is divided into two components: algorithm assessment in the context of shadow removal in diverse environments, and the creation of the adaptive model for Physical shadow removal.

We first assess the previously catalogued algorithms (Chromacity, Physical, Geometry, SRT, and LRT removal) for sensitivity to both environmental and parametric change. We perform the analysis by applying each shadow removal method to each environment represented by our datasets. The detection and discrimination rates of each method's shadow removal are plotted over time. Each method is judged based on its (detection/discrimination) response to the different environments of the datasets, and its response over time within each dataset.

We then perform a similar sensitivity analysis upon the parameters utilized by the algorithms. An algorithm is first explored through the use of interactive graphical tools, which enable exploration of a parameter's effect upon an algorithm. Parameter value ranges are also methodically swept to reveal the sensitivity of an algorithm's detection and discrimination rates to a give parameter's value.

From the sensitivity assessments, we demonstrate the creation of an adaptive model for Physical shadow removal's *coneR1* parameter. We create this adaptive model as a proof-of-concept, demonstrating the capability of correlating environmental properties to an algorithm's parameter, and automatically calibrating the algorithm for improved shadow removal. The adaptive model is created in five steps:

1. Determine optimal value of parameter (per frame): In accordance with our sensitivity analysis, each frame of a dataset has an optimal value for the selected parameter, i.e., there is an optimal value for which shadow

16

removal is maximized. The process for extracting these optimal values is detailed in section 3.1.1.

2. Quantify environmental properties: Observed environmental properties are the most influential factor for creating an adaptive model. Section 3.2 details identifying and quantifying salient environmental properties.

3. Correlate optimal parameter value and environmental properties: For each dataset, optimal parameter values and environmental properties are recorded for each frame. By analyzing the correlation between the two sets, we receive a quantitative understanding of how well that environmental parameter would serve as the basis for an adaptive model. For example, if our optimal parameter value increases by $x\%$ from frame 1 to frame 2, and our environmental property also increases by $x\%$ from frame 1 to frame 2, that environmental property may be used to predict an appropriate value of our algorithm parameter. Consistent correlations are sought across each dataset to eliminate false positives.

4. Improve/evaluate correlations: Many environmental properties exist that may not display direct correlation with a parameter's optimal value. We instead utilize these indirect properties as modulators to improve correlation observed with a primary environmental property. Multiple contributing environmental properties are combined to improve correlation and thereby shadow removal. Indirect properties are

evaluated in sections 3.2.3, 3.2.4, and 3.2.5.

5. <u>Create general model for selected parameter</u>: Incorporating direct and indirect correlative environmental properties, we construct an adaptive model capable of calculating a new value for the algorithm parameter that improves shadow removal. This model is independent of dataset, and is calculated per frame, rather than applied to every frame in a dataset. Methodology for building this model is found in section 3.3.

Chapter 3

Methodology

This chapter contains the research methodology employed while implementing a proof-of-concept of an adaptive shadow removal model. Our methodology is divided into two components: algorithm assessment in the context of shadow removal in diverse environments, and the creation of the adaptive model for Physical shadow removal, one of the previously assessed algorithms.

We first assess the previously catalogued algorithms (Chromacity, Physical, Geometry, SRT, and LRT removal) for sensitivity to both environmental and parametric change. We then perform a similar sensitivity analysis upon the parameters utilized by the algorithms. These assessments are performed qualitatively to determine the feasibility of an adaptive model being created for each algorithm.

From the sensitivity assessments, we demonstrate the creation of an adaptive model for physical shadow removal's *coneR1* parameter, which bounds

the brightness deviation used to distinguish shadow pixels from foreground. We consider various environmental properties and evaluate their correlations to the selected algorithm parameter. Potential indirect correlative factors are detailed in sections 3.2.4 and 3.2.5. Finally, methodology concerning assembling the adaptive model follows in section 3.3.

3.1 Algorithm Assessment

We qualitatively assess several leading shadow removal algorithms (Chromacity, Physical, Geometry, SRT, and LRT removal) for their sensitivity to different environmental properties. Environmental properties are identified across diverse datasets, examples including shadow darkness, shadow directionality, and color saturation. Environmental property shifts, such as illumination changes, are also observed in multiple datasets.

In addition to analyzing the algorithms' sensitivity to environmental change, we similarly assess the algorithms' sensitivity to parametric shift, i.e., algorithm performance is qualitatively assessed according to the algorithm's dependence on one or more curated parameter values. Analysis and assessment of the previously presented shadow removal algorithms is conducted using a series of graphical tools in conjunction with ground truths identifying shadow regions across eight different datasets with seven unique environments.

3.1.1 Data Collection

This section details the datasets used to conduct our assessment, as well as tools developed for analyzing the sensitivity of algorithms and parameters. The tools and techniques covered allow for both qualitative and quantitative assessment.

Datasets

A total of eight datasets were chosen for this assessment, and are summarized in Table 3.1. Datasets 2 - 8 (denoted by the prefix "aton_") are provided under the Computer Vision and Robotics Research Laboratory (CVRR) in association with the University of San Diego [**??**]. The datasets were cataloged with the Autonomous Agents for On-Scene Networked Incident Management (ATON) project [**??**]. These datasets represent diverse environments, including both indoor and outdoor scenes, a range of shadow intensity, and images of varying saturation.

In addition, two datasets were included from the Performance Evaluation of Tracking and Surveillance (PETS) project [**??**]. In contrast to the ATON datasets, these sequences were explicitly chosen because they possess subsequences featuring rapid illumination changes, due to a shift in weather conditions and cloud cover. These environmental changes darken, lighten, and blend shadows in accordance with a background model. Figure 3.1 demonstrates the illumination changes within the PETS datasets.

Name	Resolution	Sequence Length	# GT Frames
PETS1	768x576	175	12
PETS2	768x576	192	11
aton_highway1	320x240	440	8
aton_highway3	320x240	2227	7
aton_room	320x240	300	22
aton_campus	320x240	1179	53
aton_hallway	320x240	1800	13
aton_lab	320x240	887	14

Table 3.1: Dataset information.

(a)   (b)

Figure 3.1: PETS2 experiencing both high illumination (a) and low illumination (b) (due to cloud cover).

All eight datasets were manually segmented to form ground truths that identify shadow regions (gray) within foreground objects (white). Figure 3.2 illustrates the ground truths of dataset frames.
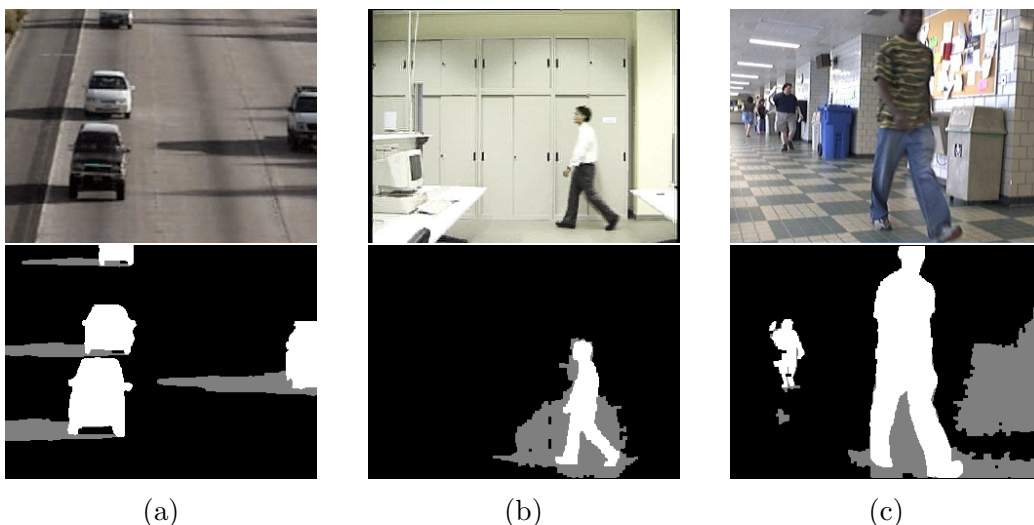
Figure 3.2: Datasets collected for ATON: (a) aton_highway1, (b) aton_lab, and (c) aton_hallway.

Assessment Tools

The first stage of our methodology is an exploration of the sensitivity of shadow removal algorithms to environmental and parametric change. We developed tools to assist intuition regarding an algorithm's performance, capable of allowing a user to manually tune an algorithm parameter, and observe its effect on performance.

To better understand an algorithm's dependence on a parameter, a framework was required to rapidly modify the parameter in question. The initial implementation of the shadow removal methods, courtesy of Sanin et al. [??], contained relevant parameters hard-coded into the algorithms. Each algorithm therefore required re-compilation to bring any modifications to fruition. This shortcoming was rectified by extracting each parameter and

Figure 3.3: An example SimpleINI file. Parameters taken from this file are used to adjust values in real-time.

organizing them into an *.ini* file. By utilizing username *brofield*'s SimpleINI architecture, hosted on Github [**??**], parameters can be either set extemporaneously or altered by an external script. An example SimpleIni .ini file is shown in Figure 3.3.

Graphical tools (GUI) were then developed to rapidly assess and visualize a parameter's affect on shadow removal. Each shadow removal method was modified to accept arbitrary parameter values in real-time from the .ini file. These parameters were then tied to graphical sliders (from OpenCV's highgui library [**??**]) dictating their range and value. Any numerical parameter that has the potential to modify the Detection or Discrimination rates of a shadow removal method is included in the GUI. An example of the GUI is seen in Figure 3.4. The process in which the GUI modifies and displays parameters is illustrated in Figure 3.5.

The display is updated in real-time with both a visual representation of
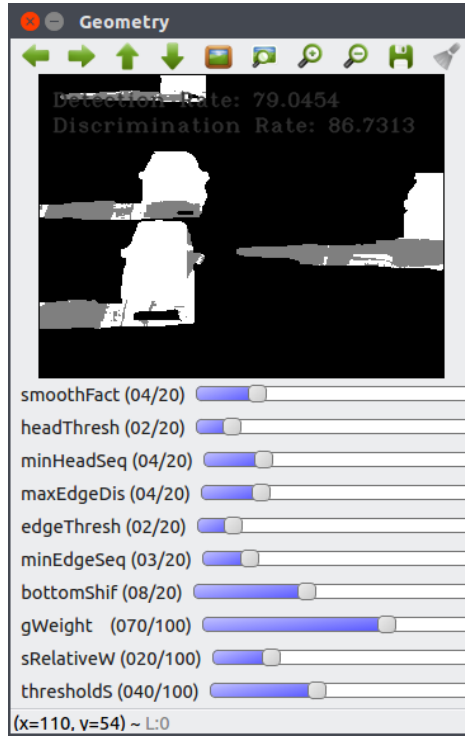
Figure 3.4: GUI tools created using OpenCV, displaying Geometry shadow removal.
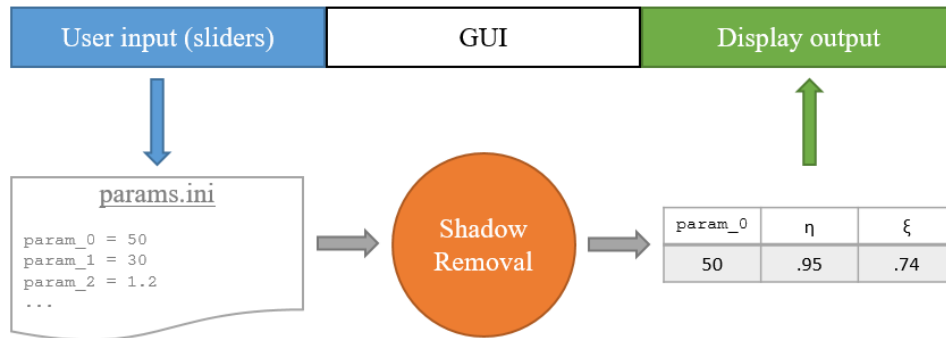
Figure 3.5: The GUI takes user input through sliders, updates values in a .ini file, which are used to produce a new output image indicating shadow versus foreground pixels with the corresponding detection and discrimination rates. The GUI then visualizes this output.

detected shadow pixels (in gray), and a quantified display of both the exact detection and discrimination rates, computed based on the ground truth. Leveraging human perception, the graphical tools enabled rapid evaluation of the sensitivity of an algorithm to changes in parameter values within specific scenes.

Determining Optimal Parameter Values

Modifying the algorithms to accept arbitrary parameter value settings at each frame enables us to create batch jobs to systematically vary parameter values across multiple runs. This allows us to meticulously sweep a given parameter (pr) through a specified range and record its corresponding affect upon detection and discrimination rates. Figure 3.6 illustrates the data collection process. The calculated detection and discrimination responses to change in pr are recorded as the functions $\eta(pr)$ and $\xi(pr)$, respectively. Both $\eta(pr)$ and $\xi(pr)$ are determined for one frame. Figure 3.7(a) demonstrates an example of these responses.

We also utilize the iterative process to determine the optimal value of pr for each frame in a dataset, i.e., which pr value yields the greatest improvements in detection and discrimination. This optimal value is denoted as pr^*. The detection and discrimination rates are summed, resulting in a shadow removal score $\Xi(pr)$ (Eqn. 3.1), which quantifies the efficacy of the shadow removal algorithm given a certain pr value. Figure 3.7(b) illustrates the score calculated using the responses found in Figure 3.7(a). We define
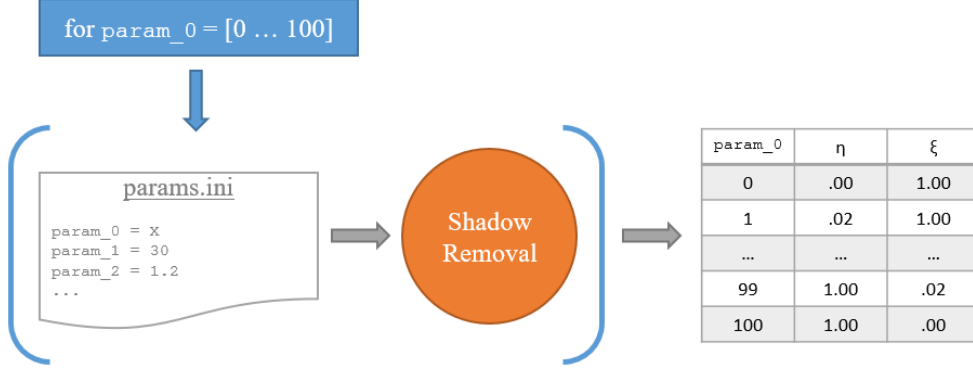
Figure 3.6: For one frame, a parameter is systematically iterated to provide detection/discrimination results for each possible parameter value.

the optimal value pr^* as the global maximum of $\Xi(pr)$ (Eqn. **??**).

$$\Xi(pr) = \eta(pr) + \xi(pr) \tag{3.1}$$

$$pr\,{}^* = max(\Xi(pr)) \tag{3.2}$$

To assess the performance of algorithms throughout a dataset, pr^* is calculated for each frame. The pr^* values are used to determine degrees of correlation for an environment, and set the foundation for a general model of arbitrary shadow removal improvement.
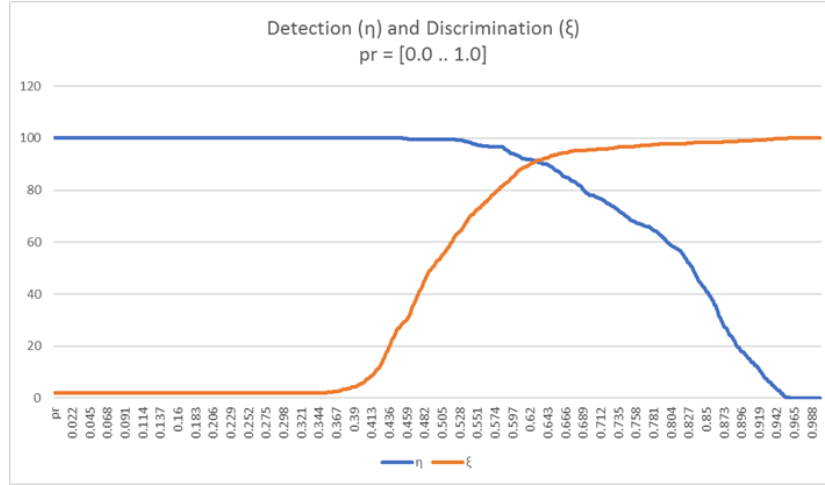
(a)
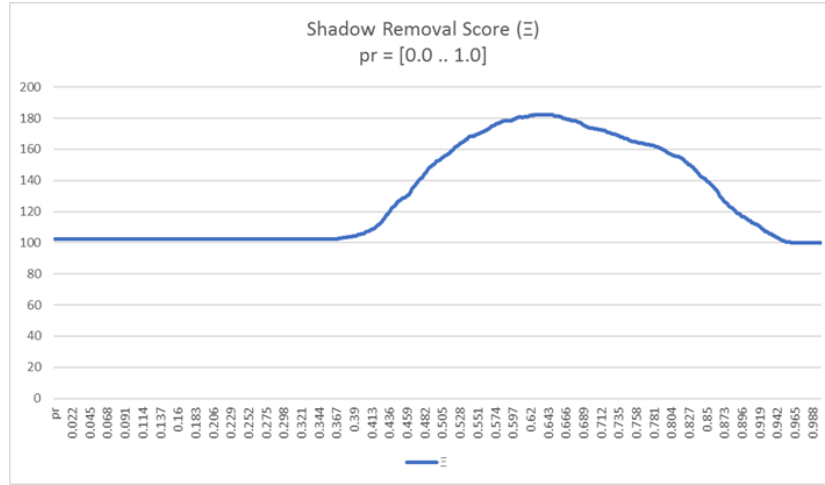
(b)

Figure 3.7: (a) Detection (blue) and Discrimination (orange) rates are charted against the iterated parameter value pr (x-axis). For each parameter value, detection and discrimination are summed to produce a removal efficacy score $\Xi(pr)$. The global maximum of this value is the optimal parameter value (pr^*).

3.1.2 Algorithm Selection Strategy

Using the tools detailed in section 3.1.1, we perform a brief qualitative assessment of the shadow removal algorithms' sensitivity. Given the diversity of shadow removal methods, a wide array of environmental factors may potentially influence an even wider array of algorithmic parameters. The scope of this research has therefore been focused on identifying a suitable algorithm to assess, linking a parameter (intrinsic to this algorithm) to unique environmental properties or changes, and developing a model for real-time parameter adaptation based on correlations found.

Candidate shadow removal algorithms are evaluated according to the correlation between their efficacy (in terms of detection and discrimination rates) over time, and qualitative observations attributed to a dataset; e.g., the PETS1 dataset experiences a large illumination change midway through its sampling, which is taken into account when searching for trends among the accuracy of a candidate method. Figure 3.8 demonstrates shadow detection rates using the PETS1 dataset, for each shadow removal algorithm.
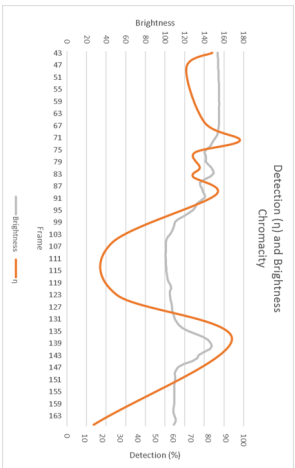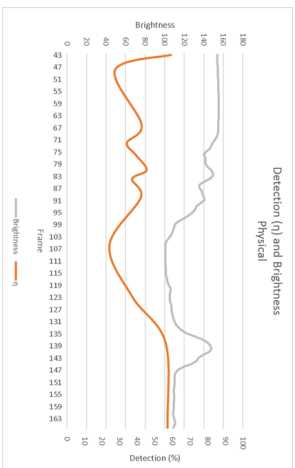
Detection (η) and Brightness
Chromacity

Brightness — η

(a) Chromacity removal.

Detection (η) and Brightness
Physical

Brightness — η

(b) Physical removal.

Detection (η) and Brightness
Geometry

Brightness — η

(c) Geometry removal.

Detection (η) and Brightness
SRT

Brightness — η

(d) Small-Region Texture removal.
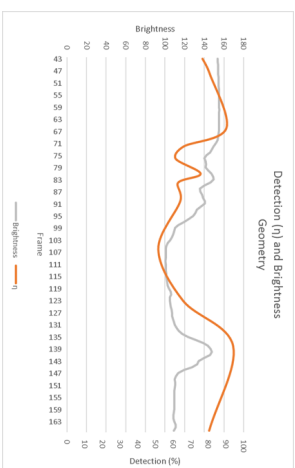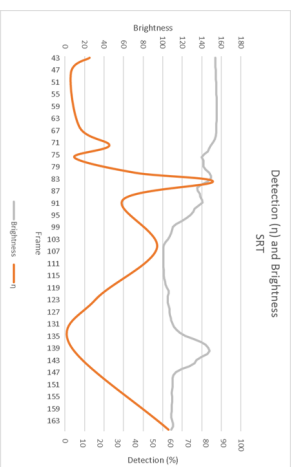
Detection (η) and Brightness
LRT
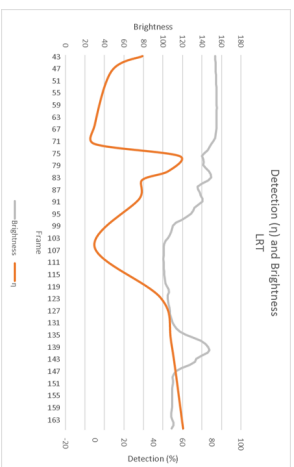
Brightness — η

(e) Large-Region Texture removal.

Figure 3.8: Shadow detection (η)(orange) for each algorithm, run on the PETS1 dataset during an illumination change. Overall brightness of a scene is shown in gray.

Evaluation of Methods

We evaluate popular shadow removal methods introduced in section 2.2: Chromacity, Physical, Geometry, SRT, and LRT. This section describes our rationale for using Physical shadow removal as part of our proof-of-concept, implemented starting in section 3.1.3. Our evaluation considers the tunable parameters of an algorithm, dependence on environmental properties and content, and sensitivity of relevant parameters.

A parameter's sensitivity is evaluated by viewing its detection and discrimination responses $(\eta(pr), \xi(pr))$. Figure 3.9 illustrates what is considered a sensitive parameter. A parameter is considered sensitive when a narrow range of parameter values affect detection and discrimination rates disproportionately. This sensitivity causes problems for an adaptive model, as the range of potential optimal parameter values (pr^*) is too small to correlate with environmental properties.

Both Geometry removal and SRT removal were shown to behave highly erratically from frame to frame even within an environmentally consistent dataset, demonstrated in Figure 3.10. Geometry-based removal remains dependent on the shape and consistency of processed foreground objects, and therefore dependent on the consistency of whichever foreground extractor is used for a scene. Furthermore, the algorithmic parameters associated with Geometry removal apply only to scenes in which the previous dependency is already fulfilled, i.e., the tunable parameters are relevant only to the ge-
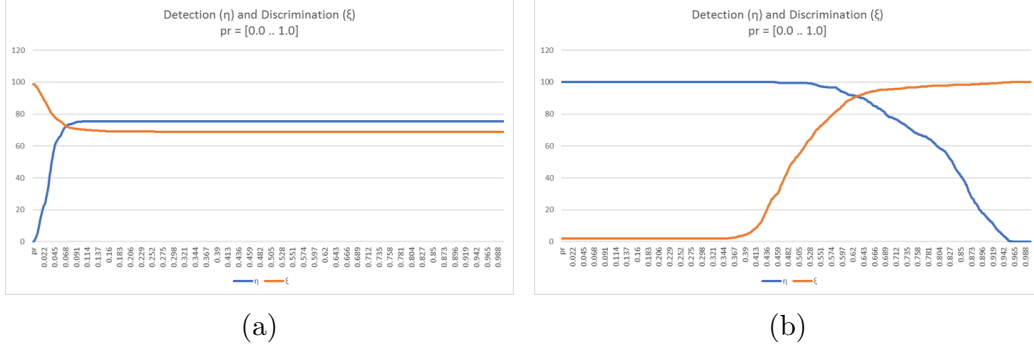
Figure 3.9: (a) depicts a sensitive parameter, while (b) represents a typical parameter.

ometric shapes necessary to enable Geometry removal. These dependencies are not related to the properties of shadows within an environment, therefore we do not pursue further analysis with Geometry removal. SRT removal is based upon a series of Gabor filters used to characterize textural patches found in shadows, and match them to a corresponding background model. This technique requires that shadows cover textural portions of the background model large enough to perform meaningful analysis. Manual tuning of SRT's algorithmic parameters provides no benefit to the algorithms detection or discrimination rates. This dependence on a shadow's area, coupled with the lack of parameters suitable to improve shadow removal, eliminated SRT removal from candidacy.

Sanin et al.'s Large-Region Texture algorithm was also eliminated from the assessment due to its inconsistent shadow removal, similar to Geometry and SRT removal (Figure 3.11). Some environments, such as the roadways of aton_highway1, yielded null detections for all frames. The Large-Region

32
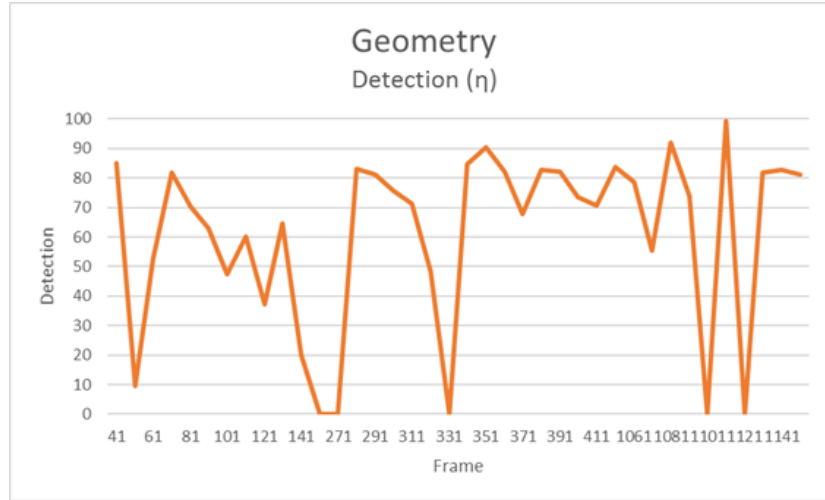
(a) Geometry
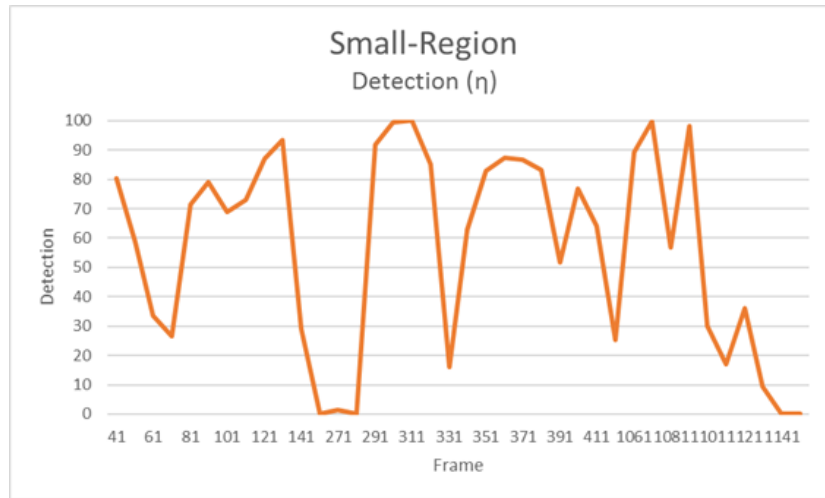
(b) Small-Region Texture

Figure 3.10: (a) and (b) showcase erratic shadow detection run on the dataset aton_campus, which experiences no significant illumination change.

Texture algorithm recognizes the pitfalls of the Small-Region approach, such as restrictive texture regions, and attempts to correct them using hard-coded parameters. The algorithm primarily displays a sensitivity to environment, as identical parameter values produce vastly disparate shadow removal performances (Figure 3.11).

LRT also demonstrates varying levels of parametric sensitivity, also dependent on the deployment environment. For example, the parameter $avgAttenThresh$ is highly sensitive, shown in Figure 3.12, and has only one value per frame that affects detection and discrimination. Furthermore, shadow removal is affected disproportionately to $avgAttenThresh$, with shadow detection rates often dropping to zero when affected. LRT has several parameters that behave in similar ways, demonstrating a lack of scalability for changing environmental properties.

Chromacity and Physical shadow removal provide tunable parameters that lie within acceptable ranges of sensitivity. An example of an acceptable range of sensitivity is seen in Figure 3.9(b). The shadow removal algorithms often respond similarly to the diversity of environments in this study; for example, Figure 3.13 illustrates comparable detection rates for the aton_hallway dataset. Chromacity and Physical shadow removal also display consistency between datasets, i.e., none of the datasets provide particularly poor results that are not linked to an environmental property change, such as illumination change. The Chromacity removal algorithm succeeds due to the tendency of shadows (within a dataset) to remain a consistent darkness when compared
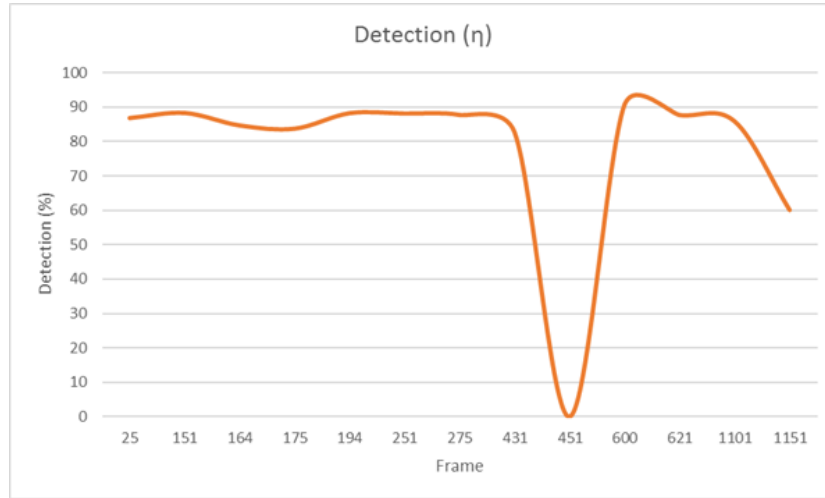
(a)
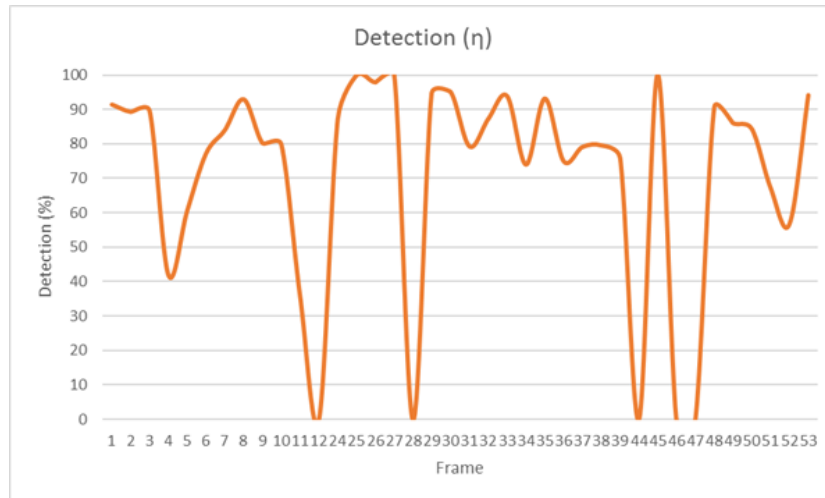
(b)

Figure 3.11: (a) Run on aton_hallway, LRT shows consistent shadow removal accuracy, with occasional dips. (b) Run on aton_campus, LRT performs erratically.

35
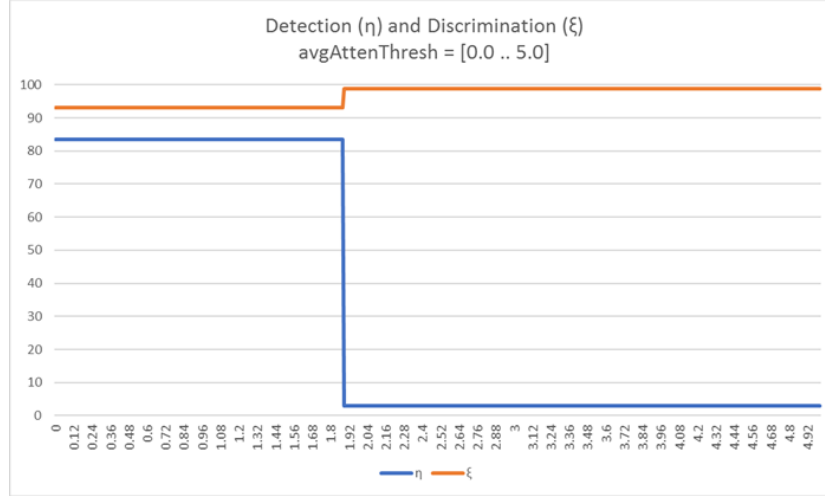
Figure 3.12: The LRT parameter *avgAttenThresh* is highly sensitive, demonstrating a narrow range for which LRT removal accuracy is affected. LRT's removal accuracy response is shown, using the aton_highway1 dataset.
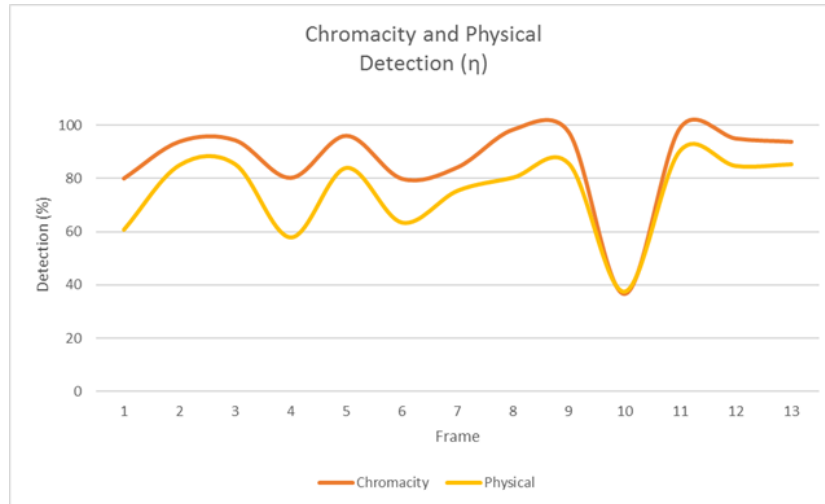
Figure 3.13: Chromacity and Physical shadow removal demonstrate similar shadow detection results run on aton_hallway.

to the corresponding background model. However, this simple approach can prove problematic when illumination changes occur within a dataset. Physical shadow removal utilizes a similar system of gauging a shadow's darkness compared to a background model - what is referred to as a weak detector [**??**] - and implicitly suffers similar breakdowns of detection when presented with significant illumination flux within a dataset. Although they both suffer, Physical shadow removal employs a strong detector [**??**] to increase detection and discrimination at the cost of processing time. In the end, Physical shadow removal was selected as the main experimental framework for this study, in part due to Physical shadow removal's more sophisticated model.

3.1.3 Selecting a Parameter of Physical Shadow Removal

Physical shadow removal operates in two stages: the weak detector, and the strong detector. The weak detector typically identifies candidate shadow pixels simply by eliminating impossible pixels, i.e., pixels that are brighter than the background model. The remaining foreground pixels are then provided as candidate pixels to the strong detector, which characterizes these pixels as either shadow or foreground. Our assessment of algorithmic parameters explores parameters in both of these detector spaces. Before we can illustrate which parameters were considered for experimentation, we must first understand each parameter's place within the algorithm.

Weak Detector - Physical Shadow Removal

The purpose of the weak detector is to prevent impossible pixels from being fed to the strong detector. The weak detector in Physical shadow removal functions similarly to the Chromacity shadow removal process; the candidate pixel is evaluated by its distance from a corresponding background pixel and sorted accordingly. The weak detector is visually, and technically, a cone projected in an RGB plane indicating the range in which a normalized shadow pixel may lie with respect to a background model. Normalizing the shadow pixel's position relative to the background model is accomplished by computing the angular distance between the RGB vector of the foreground and the RGB vector of the background. The resultant represents the color-space deviation from foreground to background. This threshold parameter is named *coneAngle* within the Physical shadow removal algorithm. The cone represents two dimensions: color deviation (θ), and intensity (or brightness) deviation. A foreground pixel is considered a possible shadow if the normalized ratio of foreground to background brightness (scaled by a function of the color deviation ($cos(\theta)$), is within a pre-defined brightness range. The tolerable brightness range, delimited by parameters named *coneR1* (minimum) and *coneR2* (maximum), is hard-coded within the original algorithm to be 0.3 to 1.0. [??] illustrates the conic model in Figure 3.14.

The main components of the weak detector are parameterized within the algorithm to fit a general model; $coneR1 < x < coneR2$ as an intensity range
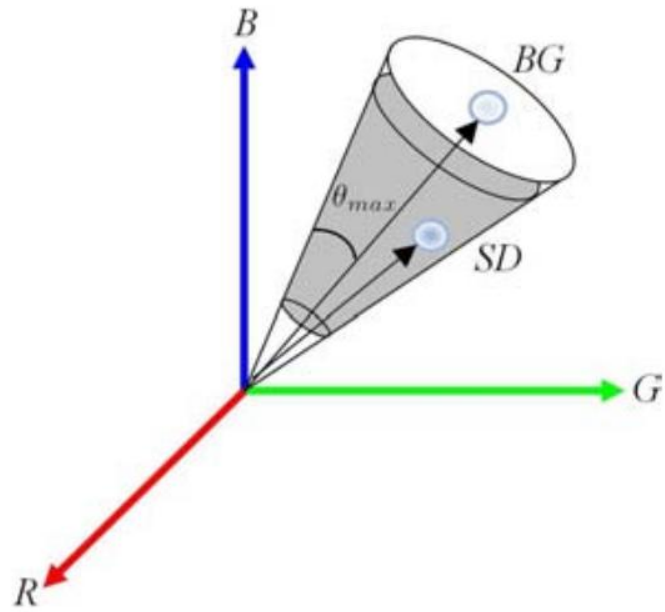
Figure 3.14: The conical volume describing the tolerable brightness range of a shadow pixel (SD), in relation to a its illuminated background pixel (BG). The maximal allowable color deviation, θ_{max}, is represented in the algorithm as *coneAngle*. Figure from [**??**].

works reasonably well for many scenes, but there exists an optimal range for any given environment. Similarly, the optimal color angle (θ) between foreground and background vectors is dependent on environmental parameters. These parameters, *coneR1*, *coneR2*, and *coneAngle* are therefore considered when looking for correlations with environmental conditions.

Strong Detector - Physical Shadow Removal

After the weak detector removes impossible candidates for shadow pixels, a Gaussian Mixture Model (GMM) is used to learn the color features of shadow pixels when compared to background pixels. Using the remaining pixels, the GMM estimates the normalized spectral ratio for shadows in a scene, i.e., the ratio of spectral illuminants [**??**, **??**]. More information on spectral illuminants can be found in section 3.2.3. The GMM is a framework based upon Expectation Maximization and learning over time; therefore, the initial parameters presented in the algorithm have little influence over the efficacy of shadow removal. Shadow removal displayed sensitivity to only one significant parameter, *postThresh*, the parameter controlling the posterior threshold. The posterior threshold is the threshold governing shadow/foreground assignment after the posterior probability of a pixel is determined via the GMM. Since the GMM adapts to its environment, modifying the posterior threshold invalidates any learning the GMM has achieved. Therefore, correlating *postThresh* to environmental properties is not performed in this study.

Evaluation of Parameters

Due to their integral nature to both the weak and strong detectors, the parameters *coneR1* and *coneAngle* are evaluated for their effect on shadow removal. Each parameter is shown to have a pronounced effect on shadow removal.

Figure 3.15 demonstrates *coneR1*'s contribution to detection/discrimination across datasets. *coneR1*'s effect on shadow removal remains consistent and scalable. The parameter is exercised through a large range, with an equally large range of resultant detection and discrimination rates. Figure 3.16 shows that each dataset demonstrates an unambiguous maximum score, *coneR1**. The wider range and relatively coarse-grained nature of the parameter facilitates identifying trends and correlations with the iterations.
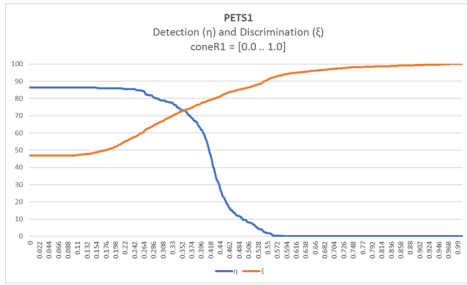
(a) PETS1
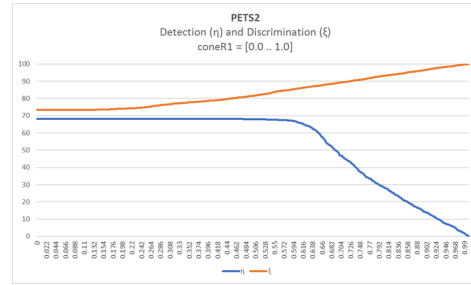
(b) PETS2
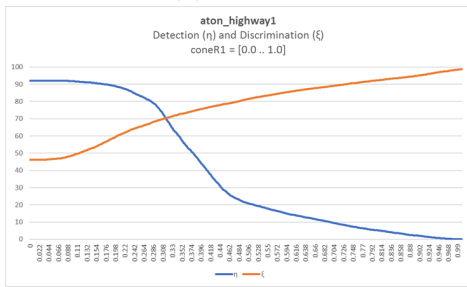
(c) aton_highway1
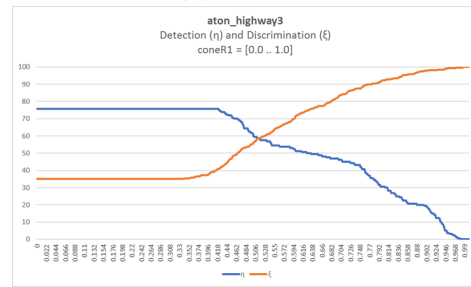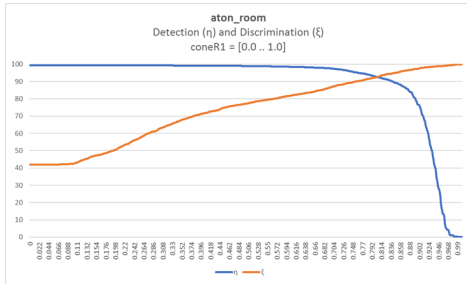
(d) aton_highway3

(e) aton_room
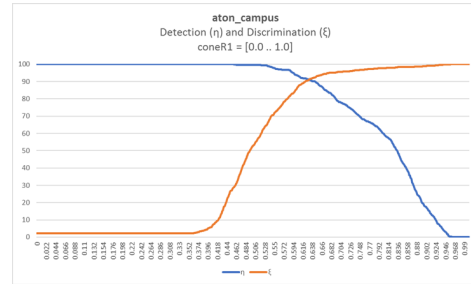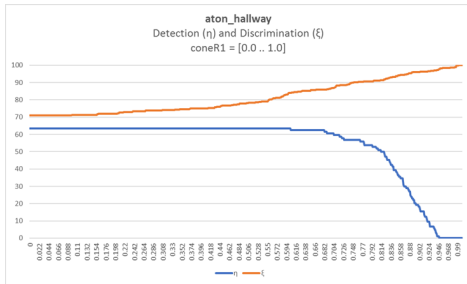
(f) aton_campus
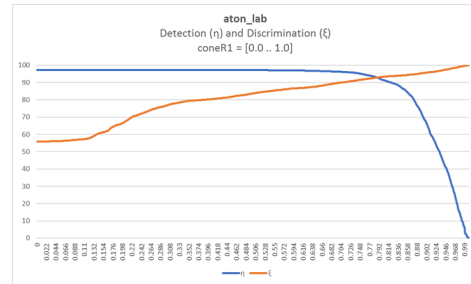
(g) aton_hallway

(h) aton_lab

Figure 3.15: Detection (blue) and discrimination (orange) rates are calculated as the value of *coneR1* is varied from [0.0 .. 1.0]. Full results found in appendices.
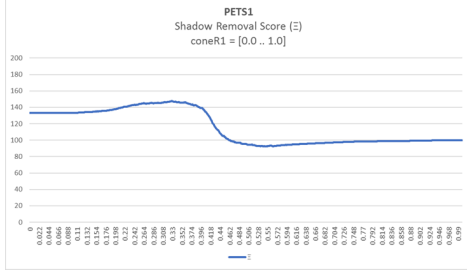
(a) PETS1
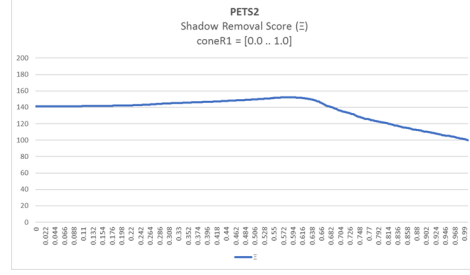
(b) PETS2
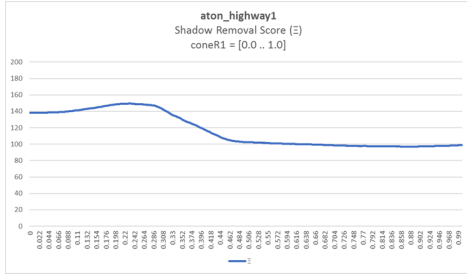
(c) aton_highway1
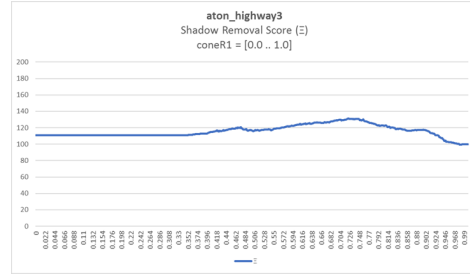
(d) aton_highway3
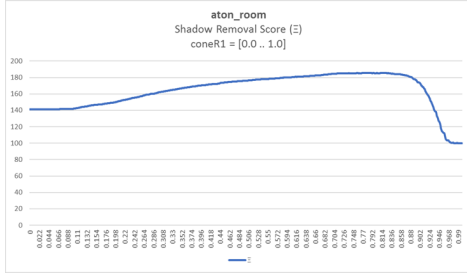
(e) aton_room
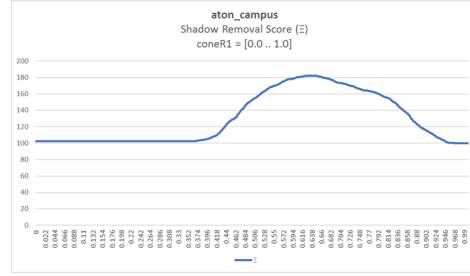
(f) aton_campus
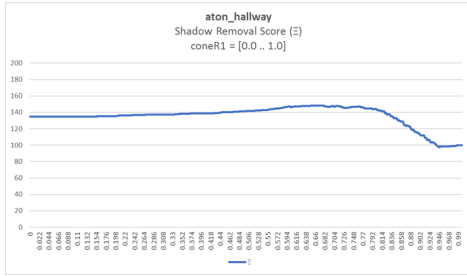
(g) aton_hallway
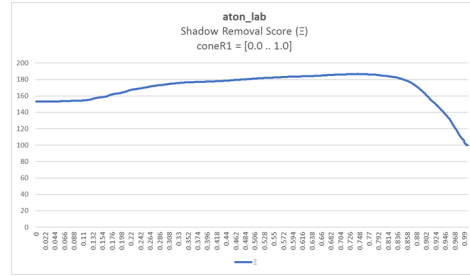
(h) aton_lab

Figure 3.16: Example score (Ξ) of a frame for each dataset. Full results found in appendices.

43

coneR2 represents the upper limit of the cone describing a shadow pixel's color representation. While it serves a similar function as *coneR1*, upon sensitivity testing, *coneR2* was never found to contribute to higher detection and discrimination rates. Its default value (1.0) is also its optimal value (*coneR2**). Figure 3.17 shows that the maximum of Ξ*(coneR2)*, *coneR2**, is 1.0 across multiple datasets.

coneAngle inhabits too narrow a range of sensitivity to properly exploit (see results included in the appendices). Figure 3.18 illustrates the sensitivity of the parameter. Within the frames of a dataset, the calculated *coneAngle** values are within millionths of one another. Furthermore, *coneAngle** does not provide as large an improvement in shadow removal as *coneR1** does. The minimal removal improvement and narrow range of sensitive values makes further examination of *coneAngle* difficult. For the duration of this study, environmental feature correlations were sought only in relation to *coneR1*.
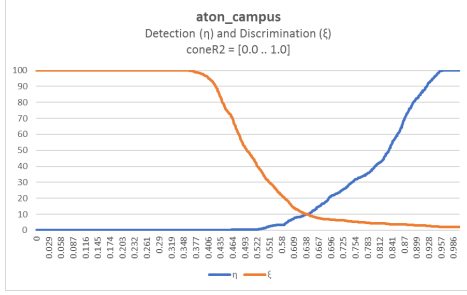
(a) aton_campus (η, ξ)
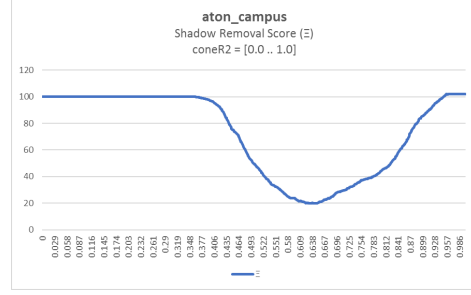
(b) aton_campus (Ξ)
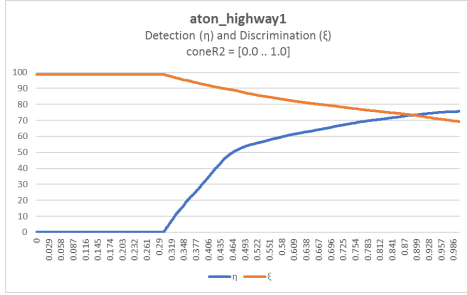
(c) aton_highway1 (η, ξ)
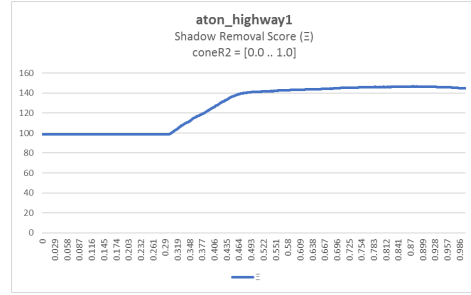
(d) aton_highway1 (Ξ)
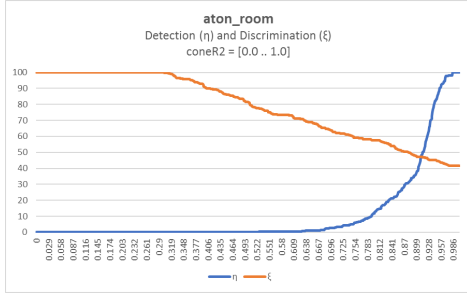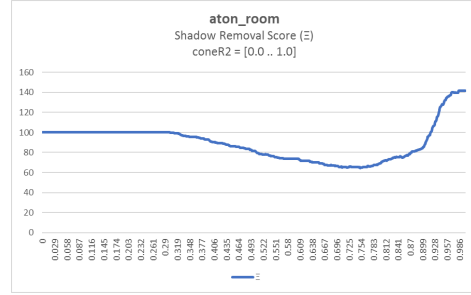
(e) aton_room (η, ξ)

(f) aton_room (Ξ)

Figure 3.17: Parameter value responses for *coneR2* for three datasets (aton_campus, aton_highway1, and aton_room). Full results found in appendices.

(a)
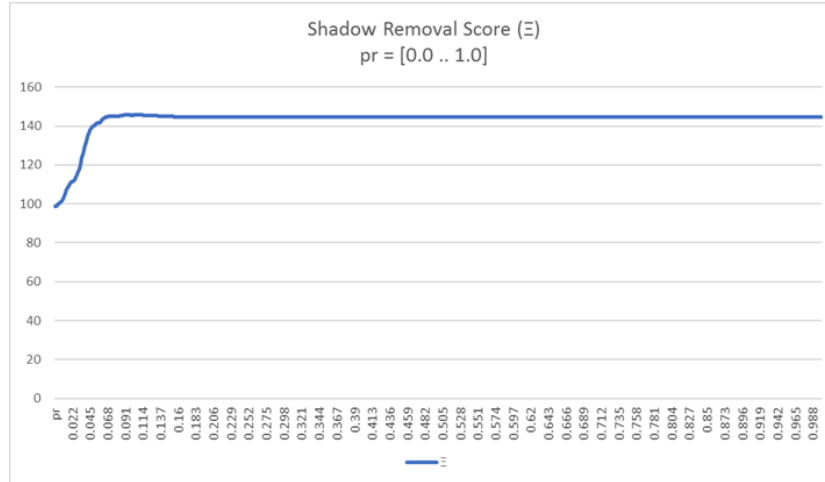
(b)

Figure 3.18: (a) Shadow removal reponse (detection and discrimination) for *coneAngle*, for aton_highway1. (b) Shadow removal score for aton_highway1. *coneAngle** is the maximum of the score.

3.2   Assessment of Environmental Properties

We now create a proof-of-concept of our adaptive model, by correlating environmental properties with a previously assessed parameter *coneR1*. We explore observed environmental properties within the datasets: attenuation and saturation. We also perform sensitivity analysis on multiple environmental properties (brightness functions, SIFT parameters) to explore their possible influence on correlation between *coneR1** and attenuation (α). A model is given for using the observed environmental properties to calculate a new variable *coneR1′*, which represents the adaptation of the *coneR1* parameter to more closely match the curated *coneR1** ideal value.

3.2.1   Previous Work - Large Region Texture Removal

Of all the leading shadow removal algorithms we evaluated, only one attempts to adapt its parameters to its current environment. In Sanin et al.'s LRT shadow removal [**??**], there are three examples of ambient environment properties taken into account: average saturation (of foreground pixels), average attenuation (from foreground to background), and average perimeter size of foreground objects. These three global properties govern the value of certain algorithm parameters.

LRT removal checks the average perimeter of foreground objects in a frame against a predetermined threshold. Average perimeter is linked to three additional algorithm parameters that control the size of the texture

region being matched to identify shadows. If the average perimeter is above the predetermined threshold, the range of area for foreground/background correlation is expanded. The average perimeter is computed on a per-frame basis, while the threshold remains static.

LRT removal, like Physical shadow removal, uses a weak detector to retrieve candidate (shadow) pixels. The candidate shadow pixels must fit within a certain range of Hue (H), Saturation (S), and Value (V)(or Brightness) when compared against the background model, i.e., the foreground HSV values subtracted from the corresponding background HSV values, referred to as ΔHSV, must fall within a specified range. There are two definitions of this acceptable range: $\Delta HSV_1 \in \{[0, 76], [0, 36], [0.6, 1.0]\}$, and $\Delta HSV_2 \in \{[0, 62], [0, 93], [0.21, 0.99]\}$. If the average saturation of all identified foreground pixels exceeds a predetermined threshold (average saturation $= 35$), ΔHS_2 ($\Delta H=[0, 62]$, $\Delta S=[0, 93]$) is used as the acceptable range. Similarly, if the average attenuation of the foreground pixels (compared to corresponding background pixels) exceeds its threshold (average attenuation $= 1.58$), the latter range of ΔV_2 ([0.21, 0.99]) is used for evaluating brightness.

This coarse-grained approach enables the LRT algorithm to switch between two presets for multiple parameters, based on observed environmental properties. However, the two presets (ΔHSV_1, ΔHSV_2), and the thresholds that govern them, require empirical knowledge about the scenes in which they are deployed. The preset-switching fails to adapt to environmental properties in a scalable manner. Applied to a 24-hour video feed, the preset-switching

model certainly fares better than the naive fully hard-coded approach, but compromises shadow removal with its rigidity. We seek to adapt parameter values on a per-frame basis, with the scalability necessary to create a continuous function of environmental properties to parameter values.

This need for adaptability underscores the motivation of this research to discover a scalable and portable solution for parameterization. In the case of Physical shadow removal, foreground object perimeter has no bearing on the accuracy of shadow removal, because Physical removal utilizes a wholly pixel-based approach. However, both average saturation of foreground objects and average attenuation are examined further in this study, as described in the next section.

3.2.2   Attenuation and Saturation

Attenuation (α) represents the loss of intensity from foreground pixel to background pixel in a frame. Traditionally, brightness attenuation is represented mathematically in decibels (dB) as the ratio of background and foreground brightness (Eqn. 3.3). The $brightness()$ function used here is the HSV brightness function (Eqn. 3.5), which defines brightness as the maximum value between the R,G,B channels of a pixel.

$$\alpha_{dB} = \frac{brightness(\vec{bg}(p))}{brightness(\vec{fg}(p))} \tag{3.3}$$

p represents pixel coordinates. In an effort to better capture shadow

pixels, we discard any pixel where $\alpha < 1.0$, implying it is brighter than its corresponding background. While the α_{dB} representation adheres most closely to the physical definition of light attenuation within a shadow, we consider the reciprocal of this model $(1/\alpha_{dB})$ of attenuation for correlation. As all considered pixels have an α_{dB} of at least 1.0, the reciprocal bounds the value between 0.0 and 1.0. This range is more suitable for correlation calculations, as the covariance between data points is easily affected by large changes in value. More information regarding correlation is found in section 4.1.

Alternatively, we consider the *percentage change* ($\%\Delta$) model of light attenuation. This model is defined in Eqn. 3.4.

$$\alpha_{\%\Delta} = \frac{brightness(\vec{bg}(p) - \vec{fg}(p))}{brightness(\vec{bg}(p))} \tag{3.4}$$

$(1 - \alpha_{\%\Delta})$ is considered for correlation, which indicates the percentage of a background pixel's intensity the foreground represents.

Saturation is the measurement of "depth of hue" in a pixel, e.g., a lower saturation means the base hue is less expressed, while a higher saturation means the color of a pixel more closely matches the base hue. The saturation level of a background pixel affects the intensity attenuation a shadow pixel experiences. Saturation properties are discussed in greater detail in section 3.2.5.
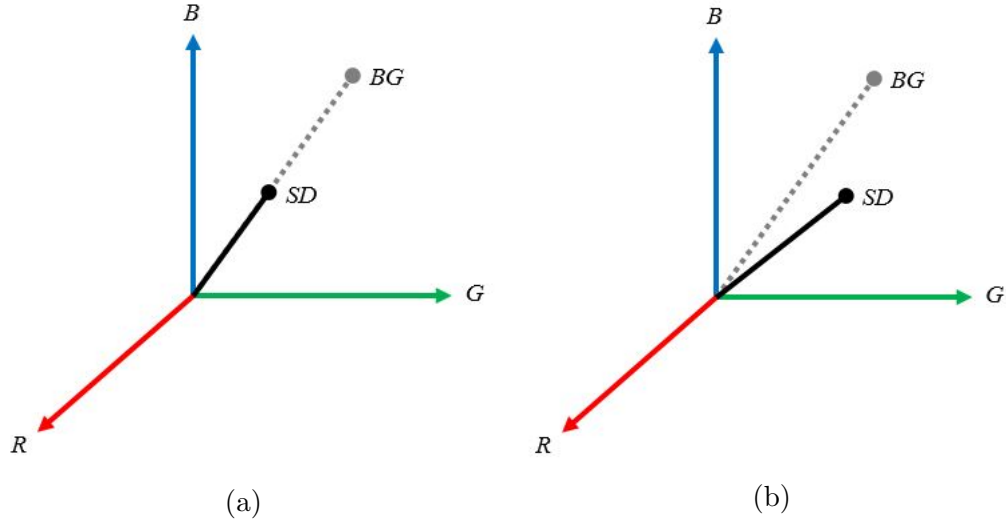
Figure 3.19: (a) Shadow pixel *SD* is attenuated linearly from background pixel *BG*. (b) *SD* is attenuated non-linearly.

3.2.3 Non-linear Attenuation and Spectral Properties of Light

In the ideal case, the attenuation of light is an entirely linear function, i.e., a shadow pixel lies along the the vector drawn from the background pixel to the origin within the three dimensional RGB space [**??**]. Figure 3.19 illustrates attenuation of a pixel linearly and non-linearly.

Due to physical properties of illumination and reflectance, this linearity does not always hold true; cast shadows in real-world scenes are caused by more than one illumination source. In an outdoor scene, a shadow may be the product of direct sunlight, blue light refracted from the sky, and diffuse light scattering from nearby objects or surfaces. Any of these factors

contribute to color bleed [**??**], influencing the attenuation of shadow pixels. These disparate light sources are said to have different *spectral power distributions*, different illumination characterized by varying concentrations of constituent wavelengths. This non-linear attenuation model is the primary motivator behind using GMM to learn a shadow's color model in [**??**].

Observed Spectral Properties in Outdoor Scenes

Since the spectral properties of an illumination source cannot always be predicted, this study attempts to observe trends across datasets to properly characterize shadows. By visualizing the magnitude of color shift due to shadows, we can clearly partition datasets into outdoor and indoor images. Outdoor datasets experience change disproportionately in each channel of an RGB image, while indoor datasets are characterized by equal and predictable shifts in each channel.

This disproportionate discrepancy is most apparent during periods of large illumination change, found prominently in datasets PETS1 and PETS2. Examining the outdoor datasets, we can see that the red and green channels experience larger magnitude shifts than in the blue channel. Based on assumed spectral power distributions of the source illumination found in the outdoor scenes, we can deduce that outdoor scenes display an introduction of blue light scattered from the sky that is not present in indoor scenes. With a spectral power distribution skewed towards blue, the red and green channels experience greater perturbation than the blue channel due to cast shadows.
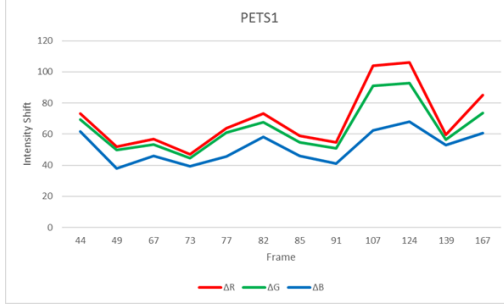
(a) PETS1
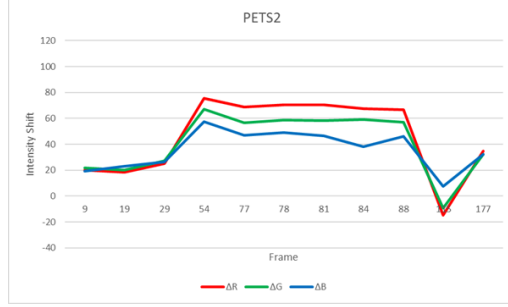
(b) PETS2
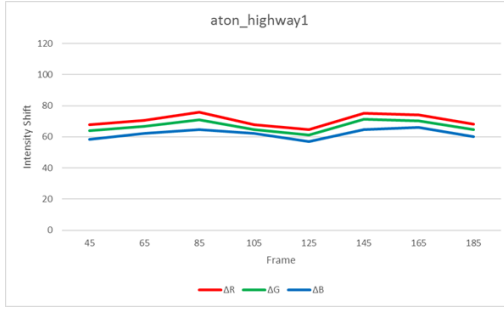
(c) aton_highway1
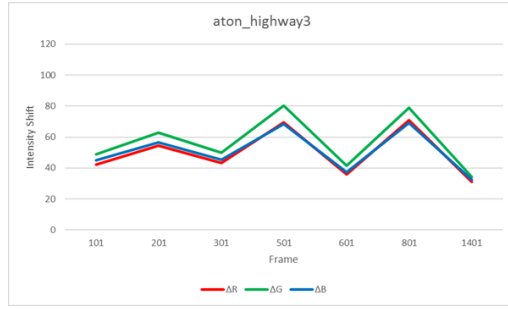
(d) aton_highway3
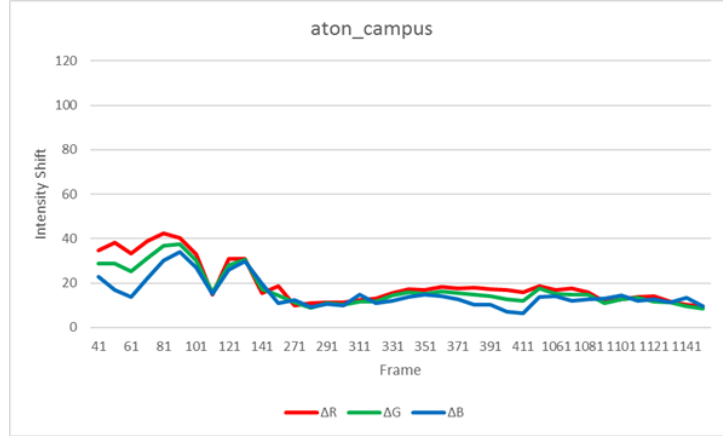
(e) aton_campus

Figure 3.20: Outdoor datasets demonstrate consistently greater deviations in the red and green channels. During illumination changes (evident in PETS1 and PETS2), the red and green channels shift disproportionate to that of the blue channel, indicating scattered blue light is a primary component of the shadows' spectral illuminant ratio.

(a) aton_room

(b) aton_hallway

(c) aton_lab

Figure 3.21: Indoor datasets demonstrate closer grouping of each channel's color shift. aton_hallway behaves most erratically of the indoor datasets. This is due to color-bleed from nearby objects, making aton_hallway the most diverse spectral ratio of the indoor sets. aton_lab behaves linearly, as expected.

This study attempts to utilize this observed multi-illuminant model to better characterize the attenuation model assumed by a scene.

3.2.4 Brightness Models

Since attenuation is a function of brightness, we attempt to utilize multiple methods of calculating brightness to improve attenuation modeling. We use different brightness functions in the calculation of α for each dataset, hypothesizing that correlation between α and *coneR1** can be improved using different brightness functions. We conduct a sensitivity analysis to determine the effect that different brightness models can have on attenuation.

Below is a short taxonomy of popular brightness models studied. Each of the brightness models included are tested and analyzed for its affect on the α model.

HSV

Described as a 'hexcone'model, [**??**], brightness in an HSV representation is defined as the maximum value of the pixel represented in the red, green, and blue channels ($\vec{RGB}(p)$).

$$V = max(R, G, B) \tag{3.5}$$

HSV can lead to misrepresentation of brightness as it is perceived by the human eye. This is because, as noted in Rec. 709/601 [**??**], the same

intensity of green appears brighter than that of red, which, in turn, appears brighter than the same intensity of blue. In the HSV scheme, green at full intensity ($R\hat{G}B = (0,1,0)$) matches the brightness of blue at full intensity ($R\hat{G}B = (0,0,1)$). It also minimizes contributions from channels other than the dominant channel, e.g., $R\hat{G}B(0,1,0) = R\hat{G}B(.9,1,.9)$. HSV therefore is best suited for environments characterized by abnormally low saturation, where light attenuation due to shadows is most linear. Environments most likely to benefit from using HSV are scenes with a single illuminant source, such as indoor scenes.

HSI

HSI represents the most basic understanding of brightness, as *Intensity*.

$$I = \frac{R + G + B}{3} \tag{3.6}$$

This understanding of brightness serves most environments properly, as it caters to each channel equally. However, HSI still suffers from the inherent luminance of certain colors, and fails to compensate for them.

HSL

HSL, a brightness representation called *Lightness*, is called a 'bi-hexcone' model [ref bihex] Lightness is an average of the primary and tertiary color compo-

nents:

$$L = \frac{max(R, G, B) + min(R, G, B)}{2} \qquad (3.7)$$

Excluding the secondary color component has the effect of translating the brightness plane described by HSI. While perceptually similar to other brightness models, HSL provides more balanced values when one channel's value is in extrema.

Relative Luminance (Y)

Originally issued in 1982, Rec. 601 [**??**] defines one of the first standards for converting analog signal into digital video. *Relative Luminance (Y)*, or, when gamma-corrected, *Luma (Y')* is the simplest extrapolation of perceptually-relevant brightness. Luminance is defined as a coefficient-weighted linear combination:

$$Y = 0.299R + 0.587G + 0.114B \qquad (3.8)$$

Rec. 709 later modified the coefficients to 0.21, 0.72, and 0.07, respectively. For this study's experimental purposes, Rec. 601 coefficients were used. This brightness model highlights the human eye's sensitivity to green hues, and is therefore particularly relevant when observing outdoor scenes.

Euclidean Norm

Taking the Euclidean norm of an $\vec{RGB}$ vector is measuring the three-dimensional distance from absolute black, $\vec{RGB} = (0, 0, 0)$.

$$Norm = \sqrt{\Delta R^2 + \Delta G^2 + \Delta B^2} \tag{3.9}$$

While the Euclidean norm does not weight the channels perceptually, as Luminance does, it does represent the most accurate way to determine the color difference between two $\vec{RGB}$ vectors, and therefore proves valuable when comparing foreground pixels to background pixels.

HSP

HSP, or *Perceived Brightness*, combines the three dimensional distance of Euclidean norm, and the perceptually weighted coefficients of Luminance. HSP was introduced by Darel Rex Finley in 2006 [**??**].

$$P = \sqrt{0.299\Delta R^2 + 0.587\Delta G^2 + 0.114\Delta B^2} \tag{3.10}$$

Ideally, HSP provides the most accurate perceptually conscious brightness. Environments that experience large saturation shifts due to shadows benefit primarily, as both color distance and weighted coefficients are considered.

3.2.5 Low-contrast SIFT Keypoints

A scene typically contains a large set of low-level textural features unique to it. These low-level features have been characterized and quantified in many ways, such as the SIFT, SURF, or FAST feature descriptors [**??**, **??**, **??**]. These feature descriptors are used for image recognition and retrieval, and are robust to varying scales, rotations, and translations. These algorithms are traditionally performed on intensity images, i.e., grayscale images. However, hue and saturation play a large role in scene characterization. Therefore, SIFT implementations were developed to incorporate color information. Popular implementations of a color-SIFT method are HSV-SIFT, RGB-SIFT, HueSIFT, and others [**??**, **??**, **??**].

Intensity images, while discarding chromatic information, often retain structural and textural information, due to image gradients' invariant basis in intensity. We assume a cast shadow similarly has minimal impact on underlying structures and textures, due to the success of textural shadow removal methods [**??**]. As a result, SIFT keypoints remain largely invariant between a frame and its background model in intensity images. Since cast shadows do not modify the underlying textural structure, they do not affect the detection of SIFT keypoints.

In a traditional understanding of shadow attenuation as a linear process, both the hue and saturation of a pixel remain constant, as the intensity attenuates in a shadow region. Since this study assumes a different, non-linear

understanding of light attenuation, we sought for representative changes that shadows bring to the hue and saturation channels of an HSV image. By detecting and displaying SIFT keypoints in the saturation channel alone, we observe qualitatively small localized structure changes within these shadowed regions (Figure 3.22).

We attempt to characterize these localized changes in structure using HSV-SIFT, by detecting SIFT keypoints using only the saturation channels of both a frame and its corresponding background model. It is apparent that foreground objects introduce significant structural changes in any of the three channels, therefore the collection of SIFT descriptors is limited to only those considered *low contrast*. The SIFT algorithm operates in four major stages: scale-space extrema extraction, elimination of low-contrast keypoints, elimination of strong edge responses, and the assignment of orientations. Finally, low-contrast detections are eliminated due to their susceptibility to noise [**??**]. In our study, we instead preserve low-contrast SIFT keypoints, to characterize the small structural changes found in shadow regions. We build two matrices: the first containing keypoints of low and normal contrast (still excluding edge responses), and the second containing only normal contrast keypoints. We then extract low-contrast keypoints using an exclusive-or operation. The non-linearity of light attenuation, paired with previously observed structural changes, make shadowed regions in the saturation channel ripe regions for low-contrast SIFT keypoints. When drawn onto the source frame, we can see that low-contrast SIFT keypoints align loosely with shad-

(a) aton_lab

(b) aton_lab ground truth

(c) Saturation channel (background)

(d) Saturation channel (frame)

Figure 3.22: The limited structural effects of cast shadows in the saturation channel (d).
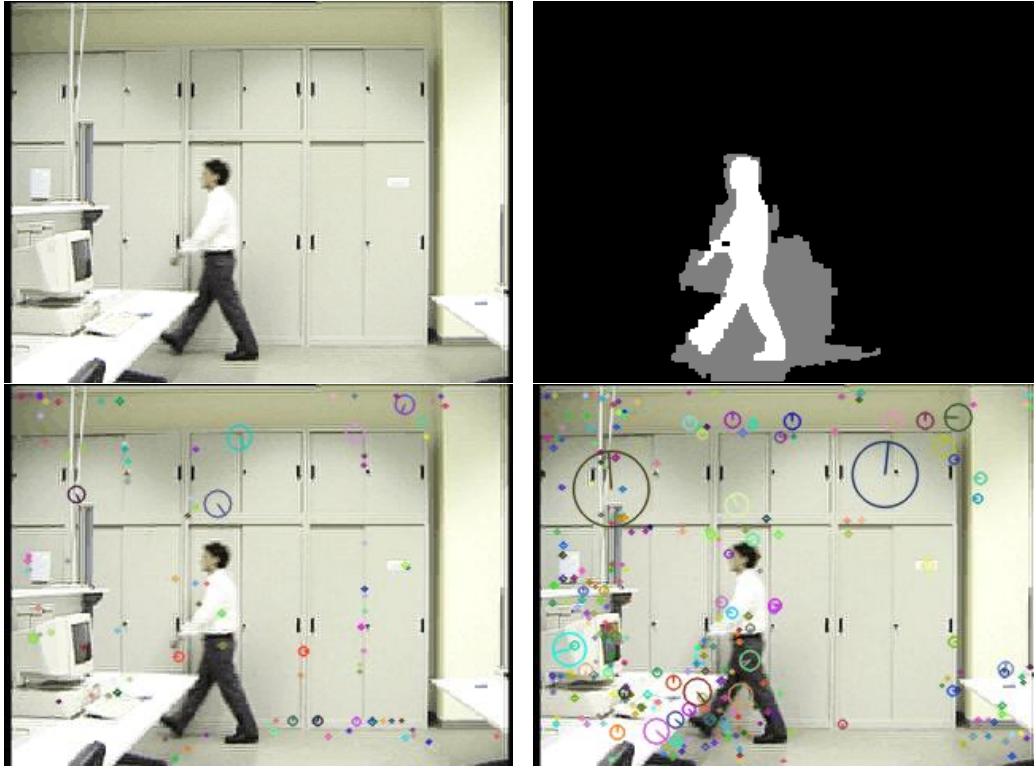
(a) Low-contrast SIFT keypoints de-   (b) Low-contrast SIFT keypoints de-
tected using the intensity channel.   tected in the saturation channel.

Figure 3.23: Detecting low-contrast SIFT keypoints in the saturation channel
more effectively captures structural changes than in the intensity channel,
due to cast shadows.

owed regions (Figure 3.23).

With shadows partially characterized by low-contrast SIFT keypoints,
this study utilizes this insight by modeling the approximate proportion of
shadows introduced in a given frame. The quantity of detected SIFT features
in a frame f is represented as $SIFT_c(f)$, where c represents the contrast ratio
parameter used. The ratio of total low contrast SIFT features to normal SIFT

features is calculated for a single frame in Eqn. 3.11.

$$SIFT_{\%C}(f) = 1 - \frac{SIFT_{0.04}(f)}{SIFT_{0.01}(f)} \tag{3.11}$$

$SIFT_{0.04}$ represents features captured using the default contrast threshold (0.04), and $SIFT_{0.01}$ represents features captured using a lower contrast threshold. We then compare the $SIFT_{\%C}(f)$ ratios of the foreground and background frames, as defined in Eqn. 3.12.

$$SIFT_{fg/bg} = \frac{SIFT_{\%C}(fg)}{SIFT_{\%C}(bg)} \tag{3.12}$$

The resultant $SIFT_{fg/bg}$ estimates the percentage change of the number of detected low contrast SIFT features from the background to the current frame. We assume a greater presence of low contrast SIFT features loosely indicates a greater quantity of shadows in a frame. In order to boost correlation between a parameter and a primary environmental property, $SIFT_{fg/bg}$ is multiplied against the value of an observed environmental property (in our case, α), serving as a scaling factor. We use this scaling factor to perform sensitivity analysis using the parameter *coneR1** and the environmental property α. Results of this analysis are found in section 4.1.2.

3.3 Weak Detector Estimation - Creating a Model

With *coneR1** calculated for a wide variety of datasets, we can begin to model a difference between environmental properties and the optimal parameter.While the attenuation from the foreground to the background displays a correlative relationship to *coneR1**, this correlation does not indicate a relationship to the magnitude of *coneR1**. This is due to the nature of the properties of attenuation. The attenuation calculation utilized is a function of brightness shift as a percentage of a background value, e.g., a pair of foreground/background pixels with a value of 50 and 100 produces the same attenuation as a pair valued at 25 and 50. Normalizing correlative environmental properties, we model an optimal shift from observed properties to the optimal value of the weak detector, i.e., the shift is modeled as $shift_{req.} = (coneR1* - \alpha)$.

$shift_{req.}$ can be extrapolated from the magnitude of color shift (ΔRGB) found from foreground to background. By plotting ΔRGB against $shift_{req.}$, we produce a general relationship. Using data points from each frame in each dataset, we can produce a best-fit polynomial to generalize a model of attenuation and brightness magnitude shift into a function dictating the required shift of attenuation (shown in Figure 3.24). This required shift is defined as $shift_{\alpha}$. To avoid over-fitting, a parametrized logarithm was chosen
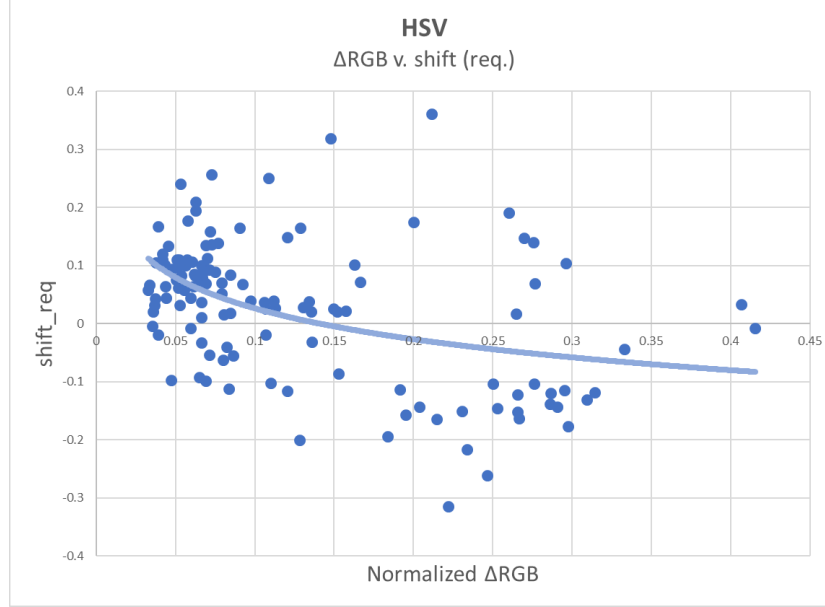
Figure 3.24: Regression performed on RGB shift found per frame, vs. required magnitude shift to optimal.

to represent the required magnitude shift (ΔRGB). The logarithm used to calculate $shift_\alpha$, derived from Figure 3.24, is defined in Eqn. 3.13.

$$shift_\alpha = -0.11307 * \ln(\Delta RGB) - 0.1884 \tag{3.13}$$

We are now able to calculate a new algorithmic parameter, $coneR1'$. $coneR1'$ is calculated per frame, based on $\alpha_{\%\Delta}$. Calculating $coneR1'$ is shown in Eqn. 3.14.

$$coneR1' = (1 - \Delta RGB) * (1 - \alpha_{\%\Delta}) + shift_\alpha \tag{3.14}$$

65

Chapter 4

Results

This chapter contains results for the correlation of algorithm parameter *coneR1** and environmental property α, improvements to this correlation, and resulting calculated parameter *coneR1′*. Section 4.1 describes the correlations (ρ_α) found between *coneR1** and both models of average attenuation (α_{dB} and $\alpha_{\%\Delta}$). The correlations are compared based on their fit to *coneR1**. Section 4.1.2 contains results regarding two indirect environmental properties ($SIFT_{fg/bg}$ and brightness calculation method) and how they affect correlations (ρ_α). The parameter *coneR1′* is calculated, and its effect on the detection and discrimination of Physical shadow removal is shown in section 4.2. The resultant detection and discrimination rates are compared to those generated by the default parameter value.

4.1 Correlation of Parameters

We observe correlation between two parameters using the statistical definition of the correlation coefficient ($\rho_{X,Y}$), defined in Eqn. 4.1.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{4.1}$$

cov represents covariance, and σ_S represents the standard deviation of a set S. The correlation coefficient is unitless, and measured from $-1.0 < \rho_{X,Y} < 1.0$. A correlation coefficient of 0 indicates the variables X and Y are uncorrelated, while -1.0 indicates the variables are perfectly inversely correlated; likewise, a correlation coefficient of 1.0 means X and Y are perfectly correlated. Correlation coefficient is used to estimate an environmental parameter's potential usefulness for calculating a new *coneR1* for use in an arbitrary environment. We represent the correlation between *coneR1** and α as ρ_α.

4.1.1 *coneR1** and Average Attenuation

Both attenuation models (α_{dB} and $\alpha_{\%\Delta}$) were considered for correlation against the *coneR1** values of a dataset. Correlation ρ_α is represented for both attenuation models as ρ_{dB} and $\rho_{\%\Delta}$ respectively. Results for both are illustrated in Figure 4.1 and Figure 4.2. Tables 4.1 and 4.2 present correlation coefficients for each dataset for α_{dB} and $\alpha_{\%\Delta}$.
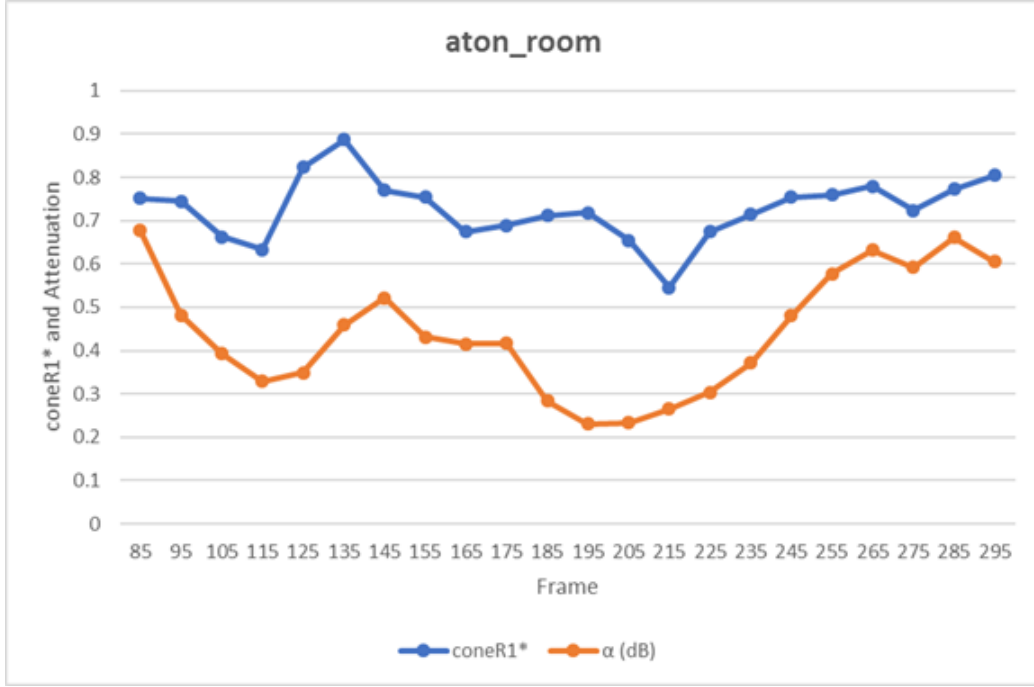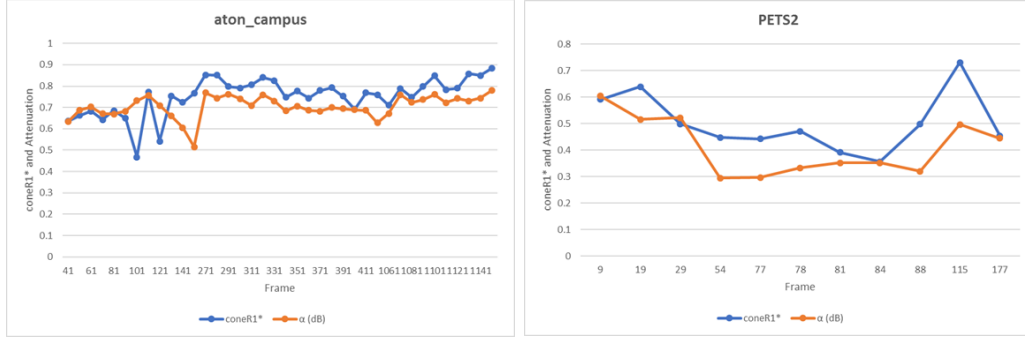
(a) aton_room

(b) aton_campus

(c) PETS2

Figure 4.1: Correlation of α_{dB} (orange) and *coneR1** (blue) is observed across all datasets. Three are shown here. (All results can be found in the appendix)
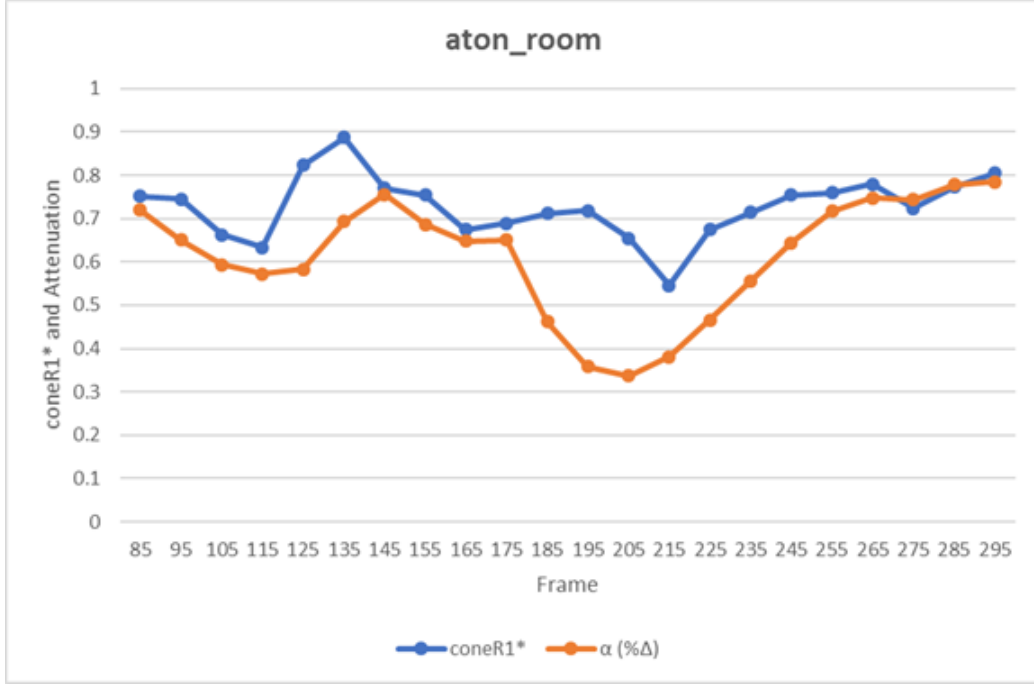
(a) aton_room
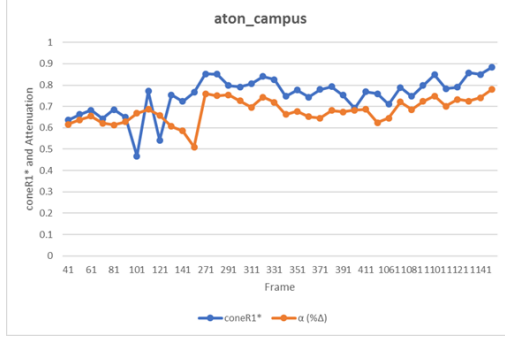
(b) aton_campus
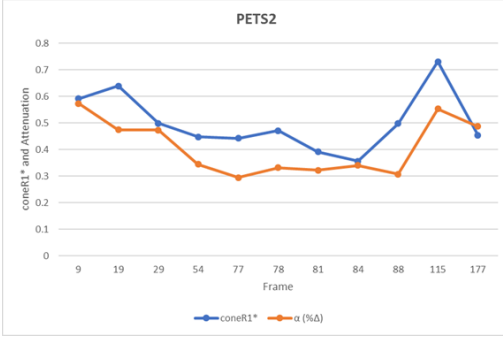
(c) PETS2

Figure 4.2: Likewise, correlation of $\alpha_{\%\Delta}$ (orange) and *coneR1** (blue) is observed across all datasets with three shown here. (All results can be found in the appendix)

Dataset	ρ_α
PETS1	0.199
PETS2	0.671
aton_highway1	0.684
aton_highway3	0.742
aton_room	0.548
aton_campus	0.360
aton_hallway	0.656
aton_lab	0.695

Table 4.1: Datasets and their α_{dB} correlations to *coneR1**.

Dataset	$\rho_{\%\Delta}$
PETS1	0.184
PETS2	0.743
aton_highway1	0.592
aton_highway3	0.801
aton_room	0.623
aton_campus	0.564
aton_hallway	0.690
aton_lab	0.551

Table 4.2: Datasets and their $\alpha_{\%\Delta}$ correlations to *coneR1**.

Dataset	ρ_{SIFT}
PETS1	-0.057
PETS2	-0.727
aton_highway1	0.009
aton_highway3	0.350
aton_room	-0.079
aton_campus	0.218
aton_hallway	-0.201
aton_lab	0.046

Table 4.3: Datasets and their $SIFT$ correlations to $coneR1^*$.

Both Tables 4.1 and 4.2 demonstrate significant correlation between $coneR1^*$ and α, consistent across all datasets. To provide contrast, an arbitrary variable (number of normal SIFT features detected) has been correlated against $coneR1^*$ (Table 4.3). In Table 4.3, we observe positive and negative correlations, as well as strong and weak correlations. Significant correlations are observed, but are inconsistent. These tables show consistent positive α correlations across datasets. Strong and consistent correlation between the two sets implies that both α_{dB} and $\alpha_{\%\Delta}$ will provide a $coneR1'$ that will improve shadow removal across the selected datasets.

PETS1 is the primary outlier with a correlation coefficient of approximately 19% for both attenuation models. While both PETS2 and PETS1 experience illumination change within the extracted samples, PETS2 and $coneR1^*$ have a correlation coefficient of approximately 74%. During the illumination change in PETS2, in which the average brightness of the entire scene decreases by 23%, observed attenuation proportionally decreases by

28%. We observe *coneR1** decreases by 37%. PETS1 experiences a similar illumination change (21%), and an attenuation change of 32%, but no significant corresponding fluctuation in *coneR1**. Illumination and parametric fluctuation was obtained by measuring these values at the beginning and end of the illumination change within the dataset, a range of 22 frames for PETS2, and 45 frames for PETS1. One possible explanation for this discrepancy is that because every pixel is darkened by roughly the same amount due to the illumination shift, the attenuation changes, while the normalized value *coneR1** does not. However, since this behavior is not evident in PETS2 as well, there are likely undiscovered factors influencing PETS1's poor correlation coefficient.

The results in Tables 4.1 and 4.2 indicate that some datasets benefit more from the α_{dB} model, while others have greater benefit from the $\alpha_{\%\Delta}$ model. While the correlation found using the two models are comparable, we use the $\alpha_{\%\Delta}$ model in Eqn. 3.14, when calculating the parameter *coneR1′*. This is because the $\alpha_{\%\Delta}$ model provides a closer fit to *coneR1**, in terms of the magnitude shift required , which is covered in detail in section 3.3.

4.1.2 Correlation Improvements

In sections 3.2.4 and 3.2.5, we discuss two indirect environmental properties that may be used to improve correlation: low-contrast SIFT keypoints and brightness calculation methods. We conduct sensitivity analysis on the two properties.

Low-contrast SIFT keypoints are analyzed within a frame and are used to produce a scaling factor that is applied to observed attenuation $\alpha_{\%\Delta}$. From this analysis we determine that the low-contrast scaling factor provides boosts to correlation in select datasets, while proving detrimental to others.

We also demonstrate that correlation (both ρ_{dB} and $\rho_{\%\Delta}$) are sensitive to varying brightness models, by contrasting their effects across datasets. We show quantifiable and predictable effects on the α_{dB} attenuation model, that indicate the HSP and Luma brightness models are most appropriate for outdoor datsets, while indoor datasets are not sensitive to varying brightness models. Similar analysis is performed regarding $\rho_{\%\Delta}$; however, the same trends (outdoor/indoor) are not observed, due to differences intrinsic to attenuation calculation.

Low-contrast SIFT Keypoints

Using Eqn. 3.12 defined in section 3.2.5, $SIFT_{fg/bg}$ is calculated per frame and multiplied against the observed average attenuation. $SIFT_{\%C}(f)$ represents the ratio of total low contrast SIFT features to normal SIFT features within a frame f.

$$SIFT_{fg/bg} = \frac{SIFT_{\%C}(fg)}{SIFT_{\%C}(bg)} \qquad (4.2)$$

The result of the multiplication is stored in the variable α_{SIFT}, and a new correlation is calculated (ρ_{SIFT}). Figure 4.3 shows the multiplication's effect
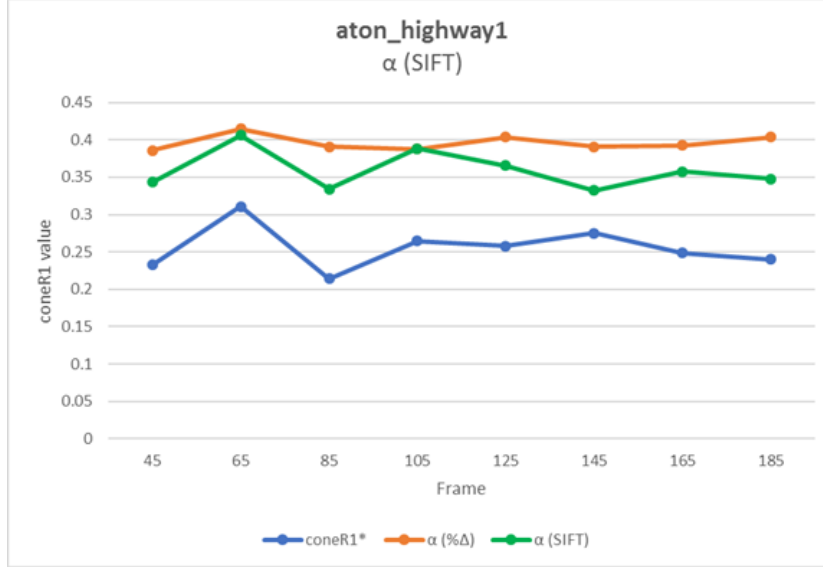
Figure 4.3: $SIFT_{fg/bg}$ multiplication's effect on correlation for the dataset aton_highway1. The observed attenuation $\alpha_{\%\Delta}$ (orange), is multiplied by the calculated $SIFT_{fg/bg}$ on a per frame basis. The resultant (α_{SIFT}), shown in green, represents a closer fit to *coneR1** (blue).

on the aton_highway1 dataset. Figure 4.4 and Table 4.4 detail the effects on correlation ($\rho_{\%\Delta}$) for each dataset.

Table 4.4 shows that modulating $\alpha_{\%\Delta}$ by $SIFT_{fg/bg}$ produces inconsistent results. For aton_highway1, aton_room, and aton_campus, the operation produced favorable results. However, for all other datasets it produced unfavorable or negligible results. The datasets that had the greatest improvements, aton_highway1 and aton_campus, contain the darkest shadows. Similarly, the most negatively affected datasets, aton_hallway and aton_lab, have the faintest shadows. From this observation, we posit that there is a range of shadow brightness in which the saturation channel is affected enough (de-

74
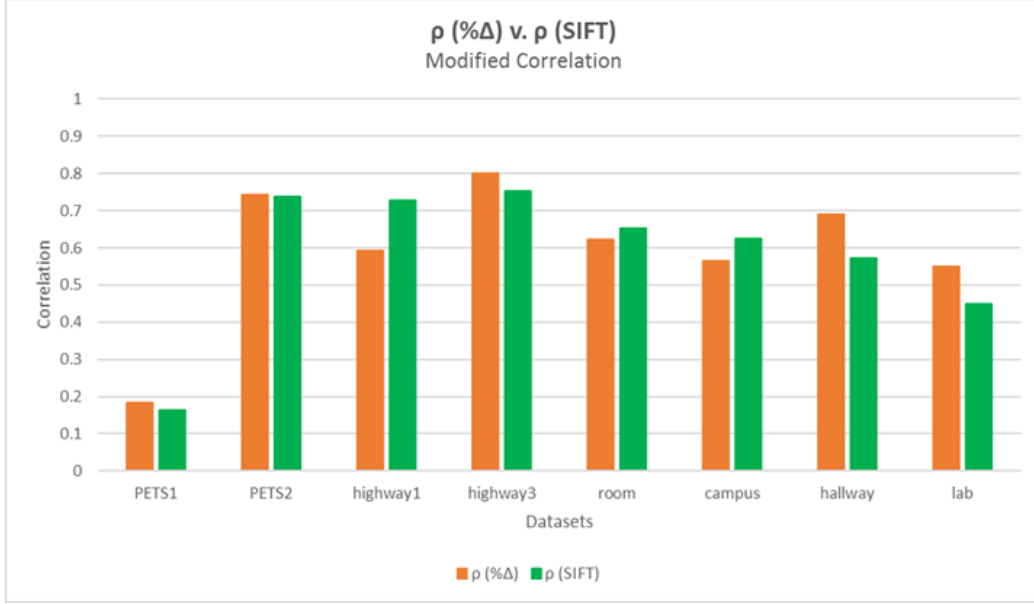
Figure 4.4: Original correlations, $\rho_{\%\Delta}$, (orange) are plotted against newly calculated correlations ρ_{SIFT} (green).

Dataset	$\rho_{\%\Delta}$ Shift (%)
PETS1	-2.06
PETS2	-0.58
aton_highway1	+13.62
aton_highway3	-4.76
aton_room	+3.087
aton_campus	+6.06
aton_hallway	-11.78
aton_lab	-10.17

Table 4.4: Correlation shifts ($\rho_{\%\Delta}$) when $SIFT_{fg/bg}$ is multiplied against observed attenuation $\alpha_{\%\Delta}$.

tailed in section 3.2.5) to be detectable as a low-contrast SIFT keypoint, and therefore that range benefits from analyzing low-contrast SIFT keypoints. This validates the assumption that low-contrast SIFT features, while not in direct correlation with *coneR1**, can provide indirect benefits by improving the existing correlation ($\rho_{\%\Delta}$), depending on the environment. Future work using low-contrast SIFT keypoints begins with determining this threshold.

Brightness Models - ρ_{dB} Attenuation Model

In addition to differing responses to attenuation models, datasets also respond uniquely to varying brightness models. Figure 4.5 illustrates how differing brightness models affect the correlation coefficient of attenuation. Table 4.5 enumerates the correlative changes experienced by each dataset when subjected to a range of brightness models. These results are illustrated in Figure 4.6.

Two distinct response trends are observed in Figure 4.6. These trends are shown in Figure 4.7: outdoor datasets (PETS1, PETS2, aton_highway3, and aton_campus) share a similar response to the various brightness models, while indoor datasets (aton_room, aton_hallway, and aton_lab) also share a similar response. Table 4.5 indicates that ρ_{dB} is sensitive to change in brightness models among outdoor datasets (Figure 4.7(a)), with HSP and Luma (Y') consistently providing the highest correlations. Indoor datasets (Figure 4.7(b)) exhibit negligible sensitivity to brightness model change for ρ_{dB} attenuation. From this indication, we deduce that utilizing either the
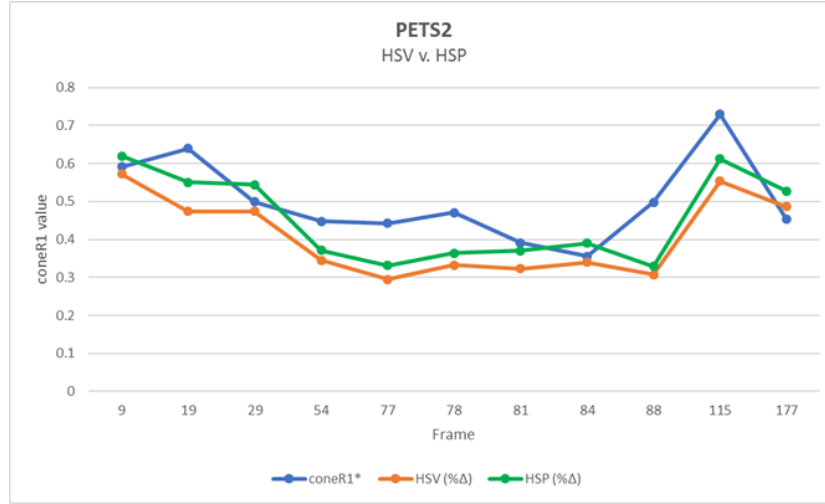
(a) PETS2
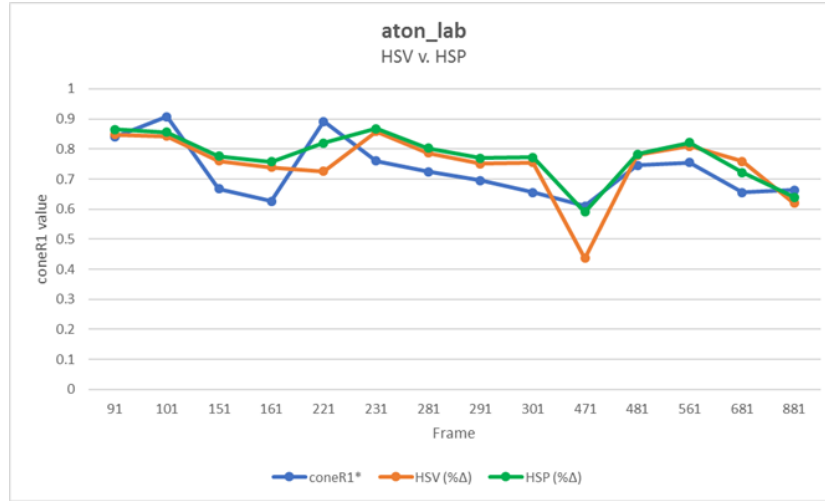
(b) aton_lab

Figure 4.5: Example contrasting correlation improvements of the HSP model (green) against HSV (orange) and *coneR1** (blue).

Figure 4.6: Datasets and their correlations (y-axis) to $coneR1*$ (α_{dB}) against Brightness models (x-axis).
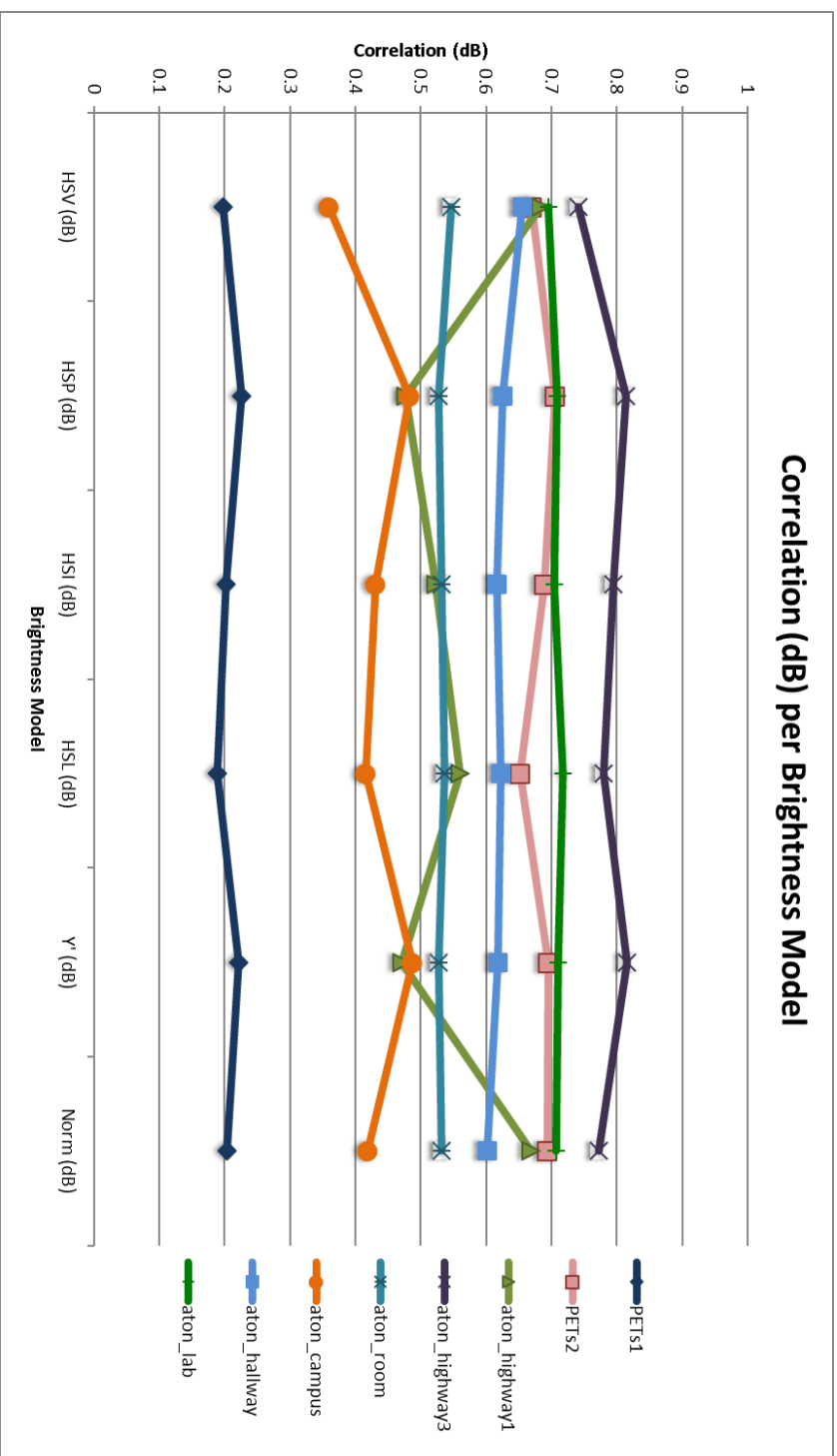
Dataset	HSV ρ_{dB}	HSP ρ_{dB}	HSI ρ_{dB}	HSL ρ_{dB}	Y' ρ_{dB}	Norm ρ_{dB}
PETS1	0.199	0.227	0.202	0.190	0.222	0.205
PETS2	0.671	0.710	0.690	0.653	0.696	0.695
aton_highway1	0.684	0.477	0.524	0.560	0.471	0.669
aton_highway3	0.742	0.814	0.795	0.780	0.816	0.773
aton_campus	0.359	0.483	0.430	0.416	0.487	0.420
aton_room	0.548	0.528	0.532	0.536	0.527	0.532
aton_hallway	0.656	0.625	0.617	0.622	0.618	0.601
aton_lab	0.695	0.710	0.705	0.717	0.710	0.708

Table 4.5: Datasets and their correlations to *coneR1** (α_{dB}) against Brightness models. Outdoor (top) and Indoor (bottom) scenes are grouped appropriately.

HSP or Luma brightness model improves correlation (ρ_{dB}) for outdoor scenes.

Furthermore, we can predict when to utilize the HSP/Luma models for better (dB) correlation, by measuring the *red-green color bias* (defined below) corresponding to shadow regions in a dataset. The difference in brightness models can be primarily characterized by their individual treatments of color content in a pixel, i.e., both Y' and HSP weight the channels of an RGB image by scaling factors relevant to human perception, while HSV simply takes the largest color channel value as the brightness. We differentiate between the datasets by measuring the average red-green color bias (β_p^{RG}, for pixel p) present in shadow pixels. To determine color bias, we use the color shift (ΔRGB) from a foreground shadow pixel to an illuminated background pixel. We define the red-green color bias as the observed blue color shift of a foreground shadow pixel p subtracted from the mean of observed red and
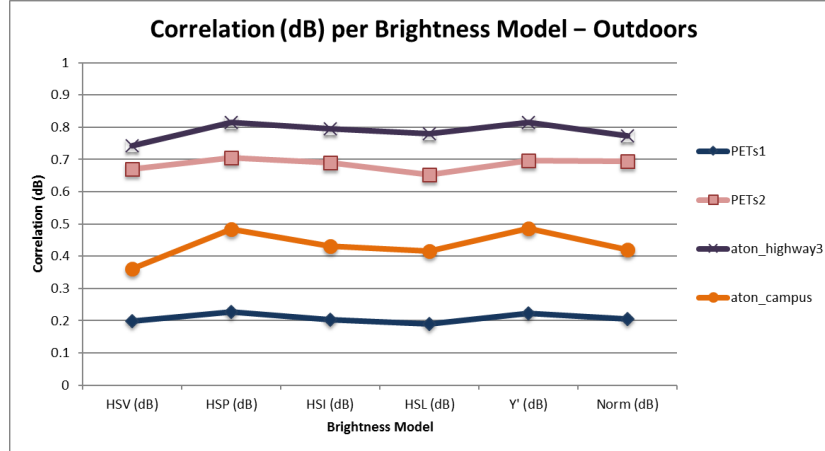
(a)
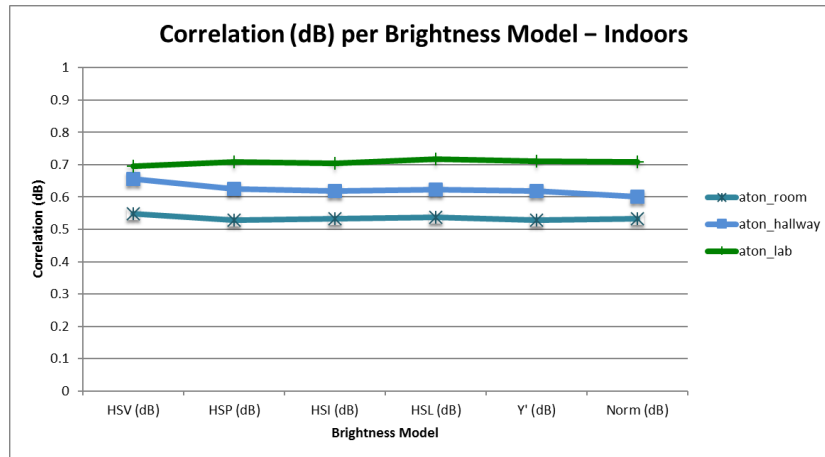
(b)

Figure 4.7: Outdoor datasets (a) share a common response to varying brightness models. In contrast, indoor datasets (b) share a insensitivity to varying brightness models

.

Dataset	β^{RG}
PETS1	15.03
PETS2	7.57
aton_highway1	6.69
aton_highway3	2.56
aton_campus	3.32
aton_room	1.64
aton_hallway	-0.26
aton_lab	0.35

Table 4.6: Red-green bias for each dataset. β^{RG} represents the average of the β_f^{RG} for each frame in a dataset. Outdoor (top) and Indoor (bottom) scenes are grouped together.

green color shifts (Eqn. 4.3).

$$\beta_p^{RG} = \frac{\Delta R_p + \Delta G_p}{2} - \Delta B_p \qquad (4.3)$$

This operation is performed on each foreground pixel that is darker than its corresponding background pixel, and averaged into β_{RG}^{f}, the average red-green color bias per frame. The values of β^{RG} are contingent on the representation of the RGB channels in an image. In our study, each channel has a range of $(0, 255)$. As seen in Table 4.6, outdoor datasets display consistently higher β^{RG} values.

The exception to this grouping of outdoor and indoor datasets is aton_highway1, an outdoor dastaset that has a response reciprocal to that of most outdoor environments. Figure 4.8 illustrates the mirrored nature of aton_highway1's response. aton_highway1 contains the darkest cast shadows in relation to its background model. The relative low points, HSP and Luma, both attempt
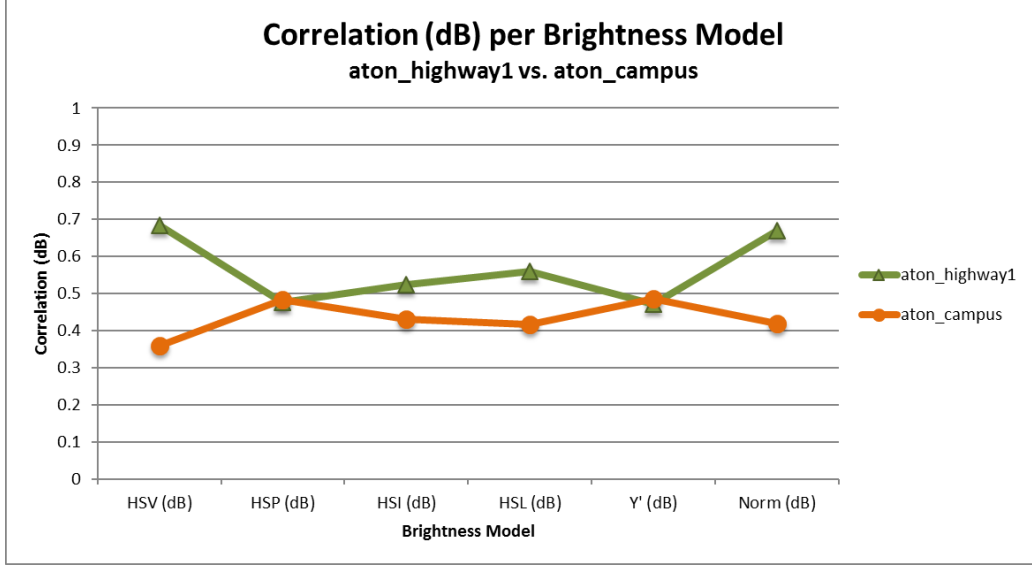
Figure 4.8: aton_highway1's response to various brightness models behaves opposite to other outdoor datasets.

to weight brightness according to color information. The applied weights are the only differing factor in the HSP and Norm methods. Therefore, we conclude that the consideration of color shift is a detriment to aton_highway1's ρ_{dB} correlations.

Brightness Models - $\rho_{\%\Delta}$ Attenuation Model

Results indicating correlations ($\rho_{\%\Delta}$) per brightness model are enumerated in Table 4.7. These results are illustrated in Figure 4.9.

It is important to note that Figure 4.9 does not exhibit the trends indicated by the α_{dB} model in Figure 4.6. The datasets are separated into outdoor/indoor datasets in Figure 4.10. Some datasets, such as PETS1, PETS2, aton_highway3, and aton_room, demonstrate similar responses to those ob-
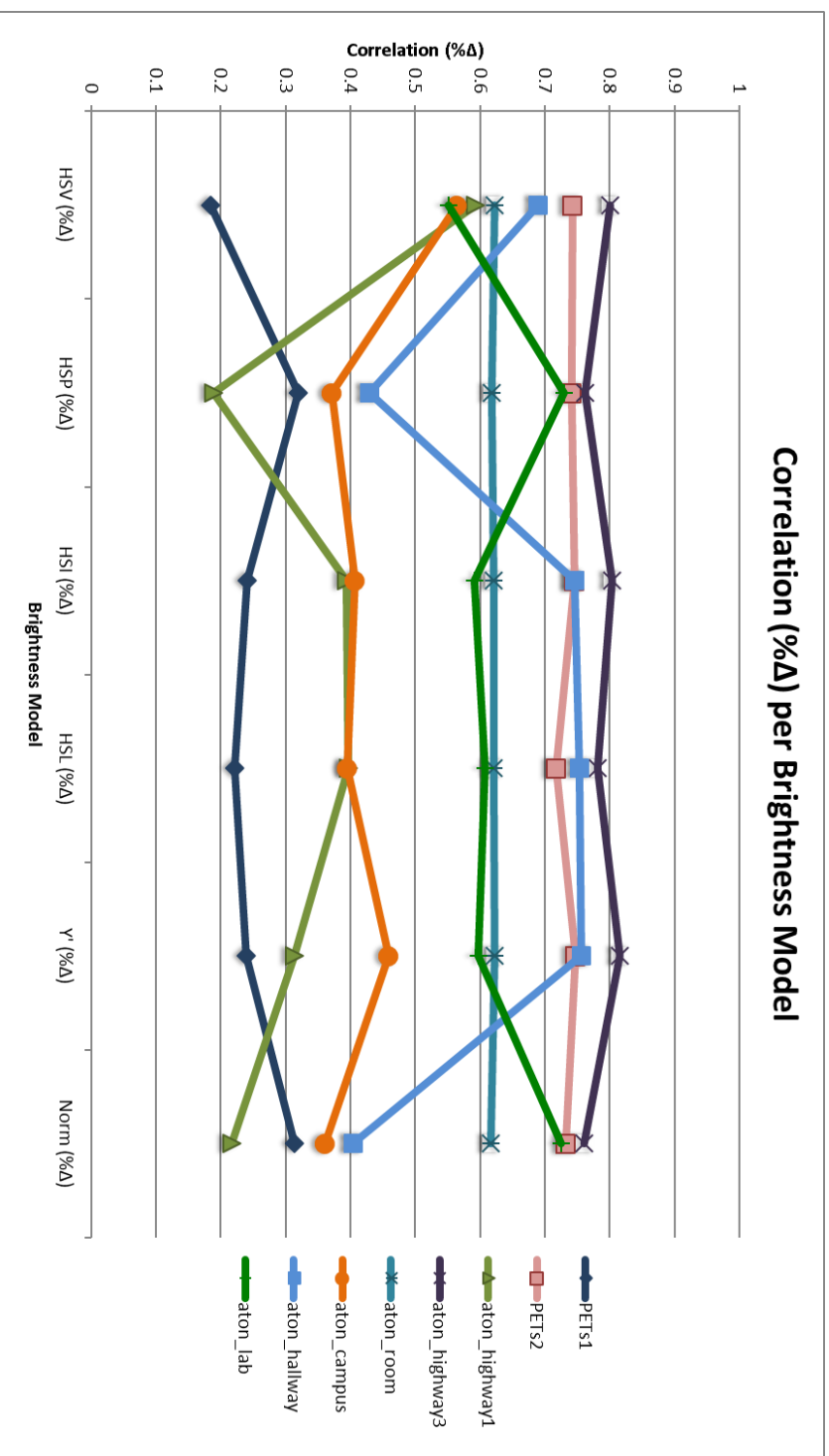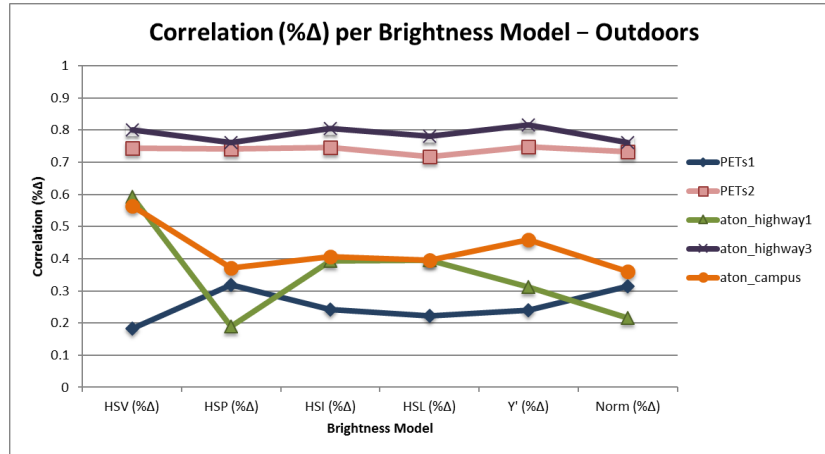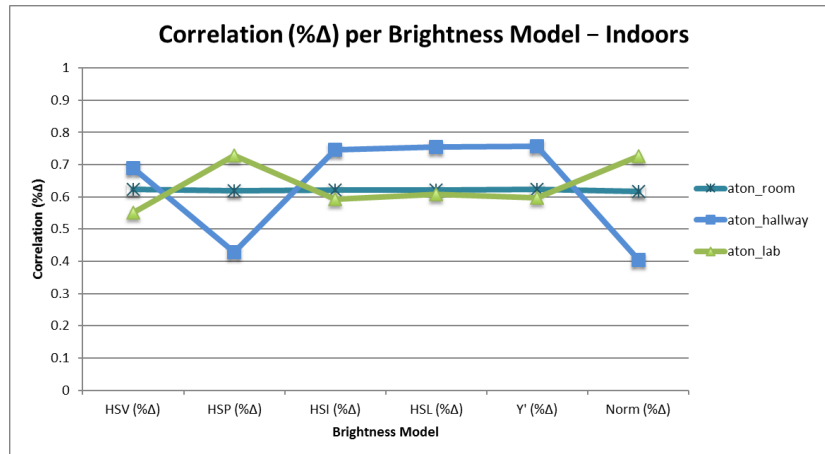
Figure 4.9: Datasets and their correlations (y-axis)to $coneR1*$ ($\alpha_{\%\Delta}$) against Brightness models (x-axis).

(a)

(b)

Figure 4.10: (a) Outdoor datasets. (b) Indoor datasets.

Dataset	HSV $\rho_{\%\Delta}$	HSP $\rho_{\%\Delta}$	HSI $\rho_{\%\Delta}$	HSL $\rho_{\%\Delta}$	Y' $\rho_{\%\Delta}$	Norm $\rho_{\%\Delta}$
PETS1	0.183	0.320	0.241	0.222	0.239	0.314
PETS2	0.743	0.741	0.746	0.717	0.747	0.732
aton_highway1	0.592	0.189	0.393	0.396	0.313	0.216
aton_highway3	0.801	0.761	0.804	0.781	0.815	0.760
aton_campus	0.564	0.371	0.407	0.395	0.459	0.361
aton_room	0.622	0.618	0.621	0.621	0.623	0.616
aton_hallway	0.689	0.429	0.746	0.754	0.756	0.404
aton_lab	0.551	0.729	0.592	0.608	0.597	0.725

Table 4.7: Datasets and their correlations to *coneR1** ($\alpha_{\%\Delta}$) against Brightness models. Outdoor (top) and Indoor (bottom) scenes are grouped appropriately.

served in Figure 4.6. However, aton_lab, aton_hallway, and aton_campus behave erratically in comparison to their corresponding α_{dB} responses. The disparity between the two brightness responses can be attributed to the vectorization of the $\alpha_{\%\Delta}$ model, which calculates the brightness of the vector between a foreground and background pixel. For example, a pixel p_1 is represented by the RGB values $(25, 50, 75)$, and a second pixel p_2 is $(75, 50, 25)$. Using the HSV brightness model, we calculate the brightness of each pixel to be 75. The α_{dB} model results in 1.0, or no attenuation. The $\alpha_{\%\Delta}$ model calculates the brightness change of the difference, $p_2 - p_1 = (50, 0, -50)$. Using the HSV model, $\alpha_{\%\Delta}$ reports a brightness shift of 50 units, with a resulting attenuation of $50/75 = 0.66$. We conclude the attenuation model $\alpha_{\%\Delta}$ is not predictably sensitive to a change in brightness model.

4.2 Parameter Model Results

The primary challenge in adapting observed average attenuation into an arbitrary model for shadow removal improvement lies in understanding the necessary translation between observed attenuation and *coneR1**. Utilizing the formulae specified in section 3.3, a coarse-grained model is developed.

Applied to an arbitrary frame, attenuation ($\alpha_{\%\Delta}$) and color magnitude shift (ΔRGB) are analyzed and used to create the necessary translation ($shift_\alpha$). This analysis is performed on each brightness model, seen in Figure 4.11.

Using ΔRGB, $\alpha_{\%\Delta}$, and $shift_\alpha$, we calculate the adapted variable *coneR1'*. Illustrated results of this model are displayed in Figure 4.12 for the datasets aton_highway1 and aton_highway3. The resultant of the model (*coneR1'*), shown in orange, retains correlative properties observed prior while providing a reasonable estimate for required translation.
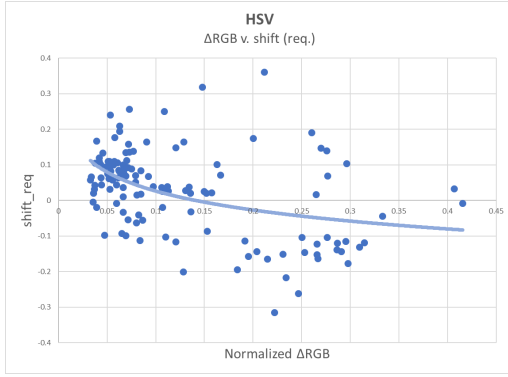
(a) HSV
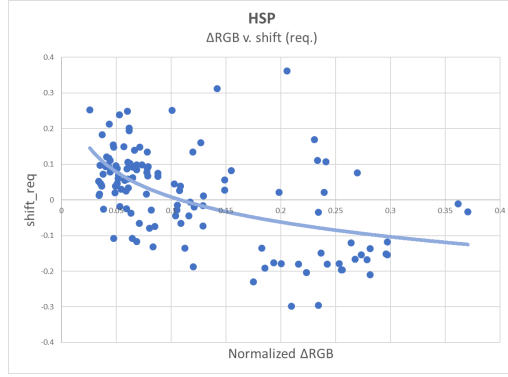
(b) HSP
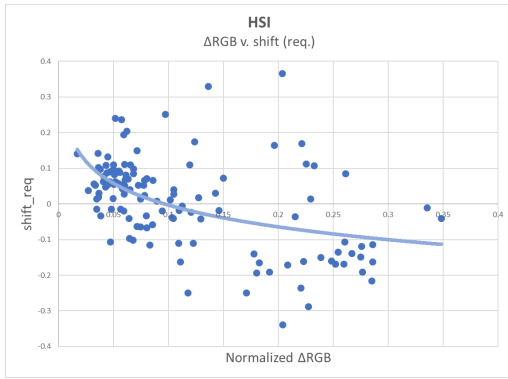
(c) HSI
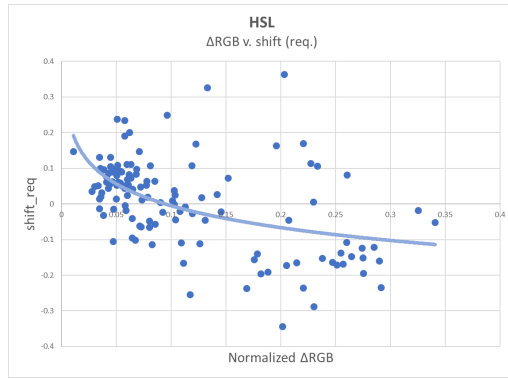
(d) HSL
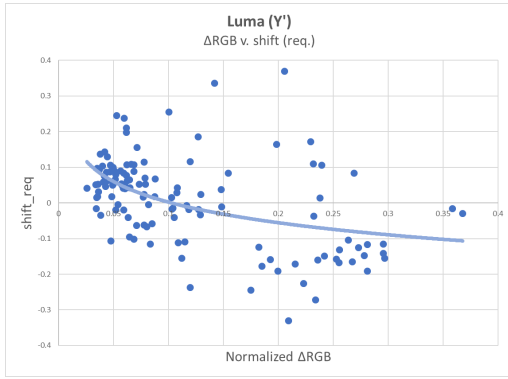
(e) Luma (Y')
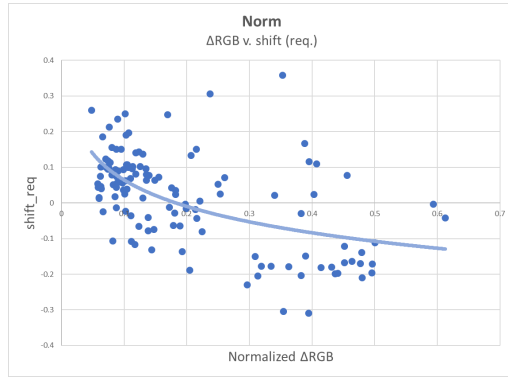
(f) Norm

Figure 4.11: For each brightness model, ΔRGB is plotted against $shift_{req.}$. The general model for each brightness calculation is formed through a best-fit logarithm.
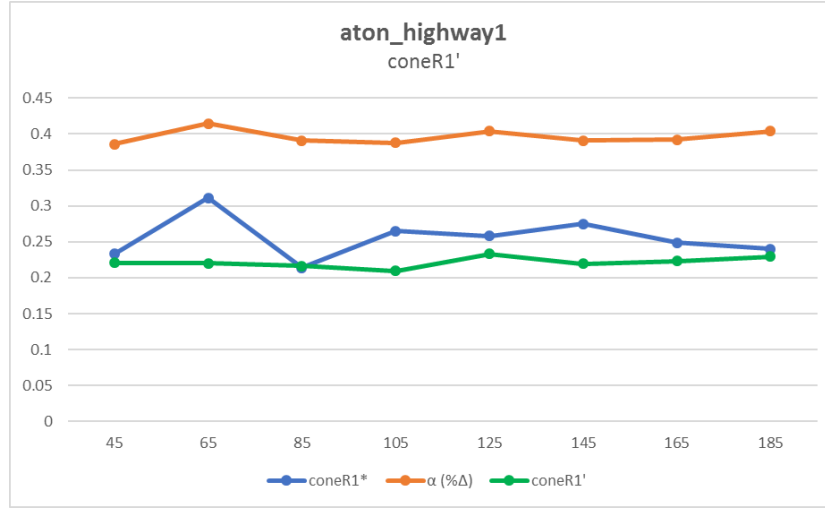
(a) aton_highway1
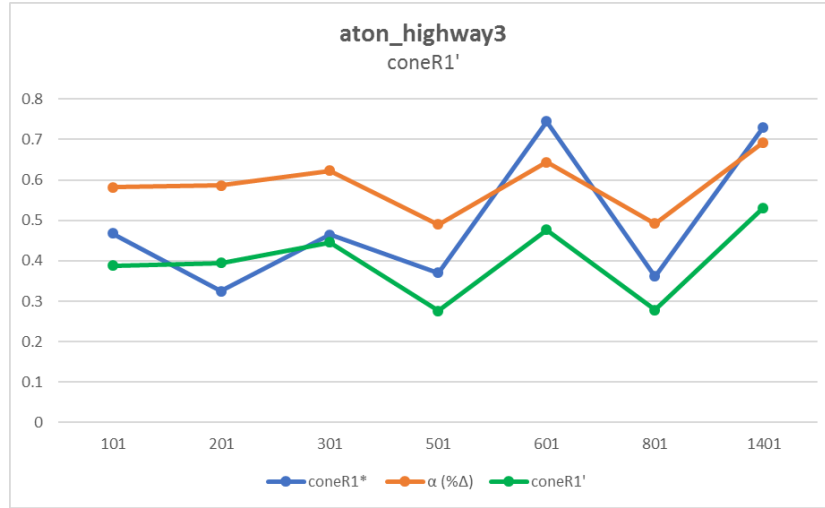
(b) aton_highway3

Figure 4.12: Adaptively tuned *coneR1′* parameter (green) charted against *coneR1** (blue) and average attenuation ($\alpha_{\%\Delta}$) (orange). All results can be found in the appendix.

4.2.1 Analysis

Figure 4.13 illustrates the resulting detection and discrimination rates of shadow removal using the adapted parameter $coneR1'$, contrasting the generated results (orange) with detection/discrimination determined via the original naive method (blue). Detection and discrimination are averaged per frame for each dataset. Quantitative results are shown in Tables 4.8 and 4.9.

Results indicate that for a majority of datasets, small amounts of shadow detection accuracy is sacrificed for disproportionate increases in shadow discrimination. Figure 4.14 qualitatively illustrates shadow removal improvements using the adaptive method for datasets aton_campus and aton_room.

aton_highway1 proved anomalous, trading in nearly equal part detection for discrimination, arriving at a net modest positive of 1.2%. We hypothesize that is because aton_highway1 has the darkest cast shadows. Furthermore, aton_highway1 attenuates non-linearly (as expected of an outdoor scene), yet appears desaturated. The combination of the strong shadows and the low saturation of the frame allows a larger population of candidate shadows pixels to be gathered. Due to these conditions, foreground *object* pixels are more likely to appear identical to shadow pixels, in both color information and relative distance from its background pixel. The confluence of these scene characteristics renders aton_highway1 less predictable than other datasets.

PETS1 and PETS2 also do not conform to the upward trends seen in the majority of datasets. In the case of PETS1, the low correlation coeffi-
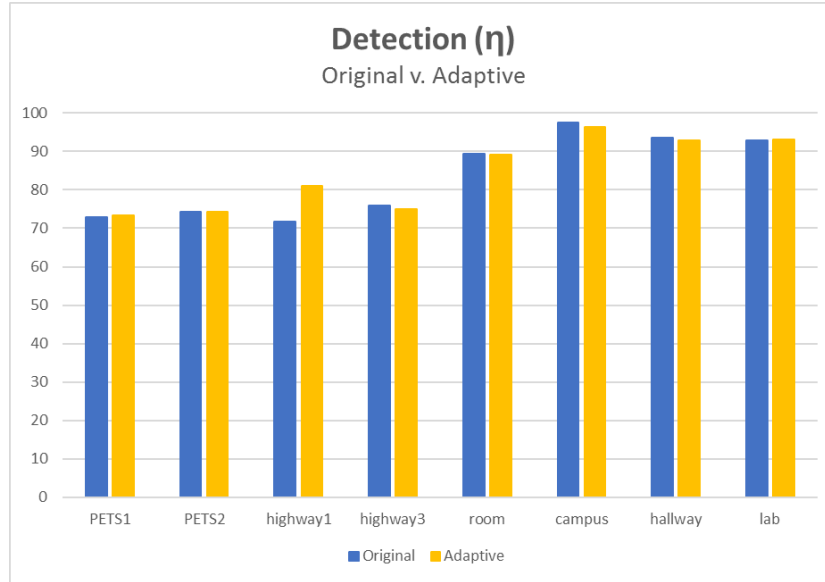
(a)
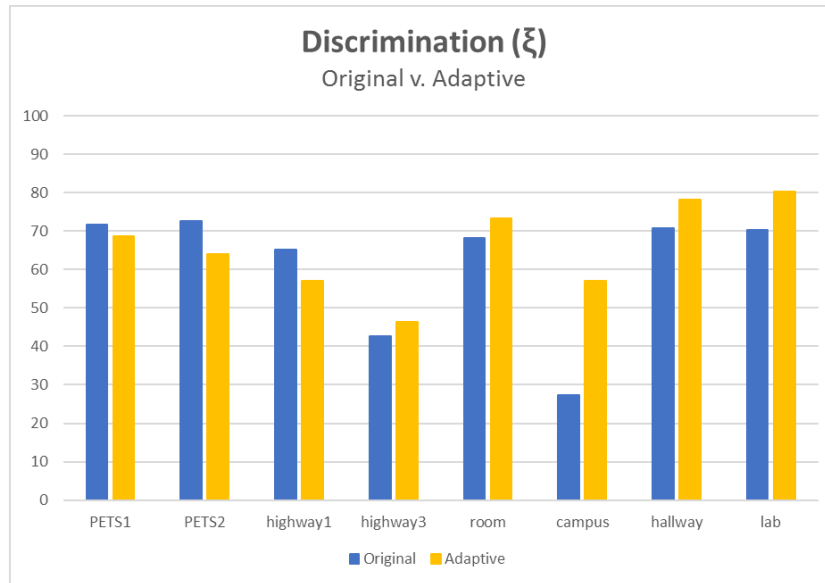
(b)

Figure 4.13: Average detection (a) and average discrimination (b) calculated using the adaptive parameter model, for each dataset.

Dataset	η (%)	+/- (%)
PETS1	73.28	+0.459
PETS2	74.34	+0.056
aton_highway1	80.99	+9.353
aton_highway3	74.88	-1.154
aton_room	89.10	-0.264
aton_campus	96.46	-0.976
aton_hallway	92.91	-0.660
aton_lab	93.12	+0.192

Table 4.8: Average detection (η) calculated from the adapted *coneR1'* (left). η is compared against original naive detection (*coneR1* = 0.3). The difference (right) is represented as a percentage.

Dataset	ξ (%)	+/- (%)
PETS1	68.61	-3.176
PETS2	64.04	-8.458
aton_highway1	57.05	-8.172
aton_highway3	46.35	+3.802
aton_room	73.36	+5.278
aton_campus	57.05	+29.67
aton_hallway	78.22	+7.549
aton_lab	80.16	+9.787

Table 4.9: Average discrimination (ξ) calculated from the adapted *coneR1'* (left). ξ is compared against original naive discrimination (*coneR1* = 0.3). The difference (right) is represented as a percentage.

(a) Ground Truth

(b) Default (0.3)

(c) *coneR1′*

(d) Ground Truth

(e) Default (0.3)
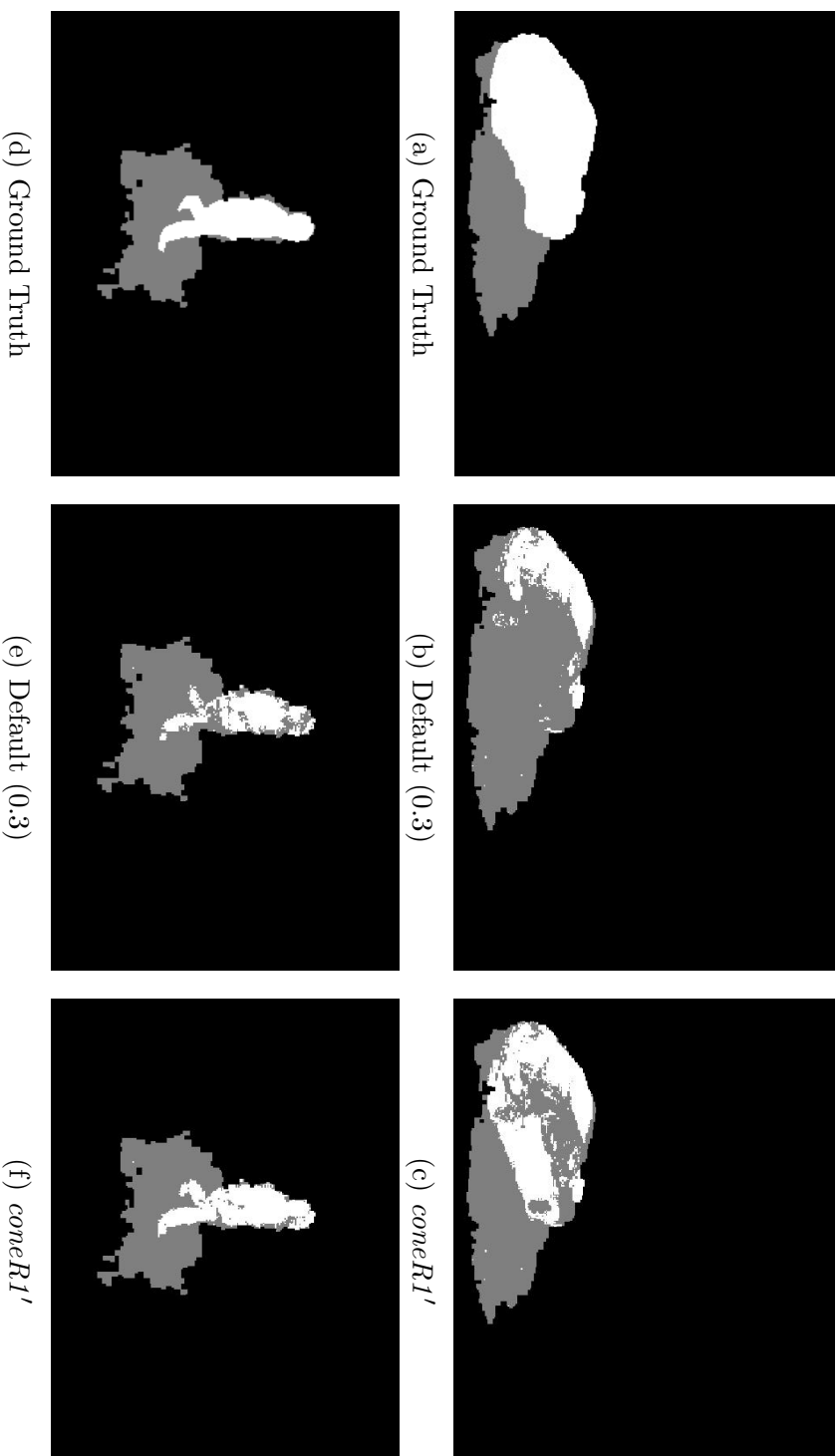
(f) *coneR1′*

Figure 4.14: Qualitative shadow removal improvement. (a - c) feature the aton_campus dataset, and (d - f) feature the aton_room datasets.
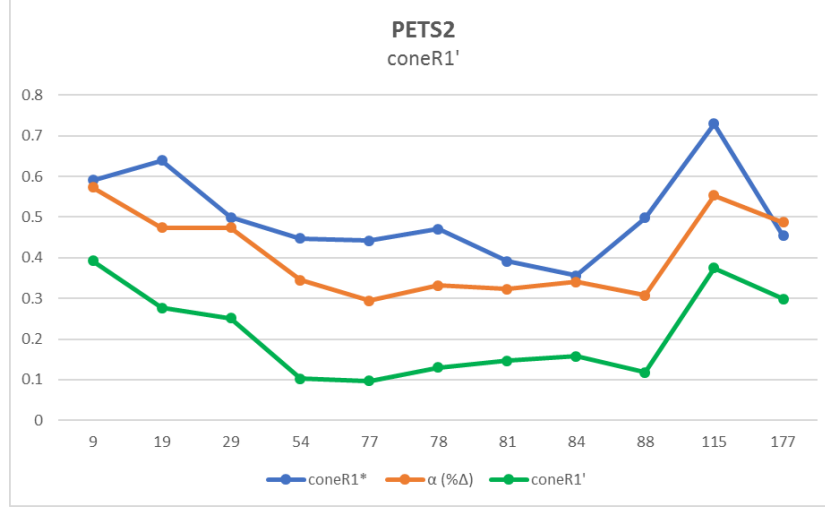
Figure 4.15: For PETS2, the adapted parameter *coneR1'* (green) is erroneously shifted downwards. The originally observed attenuation ($\alpha_{\%\Delta}$) is shown in orange.

cient infers the poor detection/discrimination observed. More interestingly, PETS2 experiences the same degradation of performance as PETS1, but has a much stronger correlation coefficient, seen in Table 4.2. This decoupling of correlation and accuracy stems from the necessary translation of the observed average attenuation by the generalized adaptation model. This is illustrated in Figure 4.15. The model overcompensates for observed color shift (ΔRGB) in PETS2, because of the illumination change in the dataset. The misrepresentation of attenuation (and thereby ΔRGB) of cast shadows in PETS2 results in the unnecessary 'downward'translation, impacting PETS2's shadow discrimination accuracy. This may be solved using techniques for rapid illumination change compensation [**??**], explained in section 5.1.

Chapter 5

Conclusion and Future Work

We have demonstrated the capability of an adaptive model to intelligently adapt the parameters of a shadow removal algorithm, using observed environmental features. This was completed through the construction of a proof-of-concept which adapted the parameter *coneR1* of the Physical shadow removal algorithm, according to the average brightness attenuation (α) observed between a shadow region and its corresponding background. The adaptive model was shown to positively affect the discrimination (ξ) of foreground object pixels from shadow pixels for 6 out of the 8 datasets, with average improvements ranging from 4% to 29% greater accuracy (when compared to the naive model). The adaptive model was shown to behave poorly with regards to rapid illumination change (PETS1, PETS2), decreasing the shadow discrimination by 3% and 8% during these periods. The adaptive model marginally increases shadow detection (η) in three datasets (PETS1,

94

PETS2, aton_highway1) by 0.5% to 9%, while the remaining datasets trade fractional amounts ($< 0.5\%$) of shadow detection for the previously observed increases in discrimination.

We have developed tools to facilitate the construction of the adaptive model, including a graphical interface for rapidly modifying a shadow removal algorithm's parameters, and an iterative process designed to automatically calculate the optimal value of a given parameter, i.e., the value that yields the highest combination of shadow detection and discrimination for a frame. Using this framework, we provided the ability to quickly assess an algorithm's sensitivity to its mutable parameters, both qualitatively, using the interactive graphical tools, and quantitatively, using the iterative process.

In order to construct the adaptive model for Physical shadow removal (and *coneR1*), we correlated shadow brightness attenuation (α) to the optimal parameter value *coneR1**, determined empirically through the interactive parameter exploration. The correlation (ρ_α) ranges from 18% to 80%. We improve correlation by evaluating two models of attenuation, α_{dB} and $\alpha_{\%\Delta}$. The adaptive model is based on the $\alpha_{\%\Delta}$ model, because it provides a better fit to *coneR1**.

We also presented two methods to affect the correlation ($\rho_{\%\Delta}$). The first of these, the measurement of low-contrast SIFT features in a scene, demonstrates changes in correlation (ρ_α) ranging from -12% to 13%. Varying the brightness model (HSV, HSP, HSI, HSL, Y', and Norm) used in calculating attenuation produced a wide range of correlation changes for both $\alpha_{\%\Delta}$ and

95

α_{dB}. For α_{dB}, we demonstrate the ability to select the best-suited brightness model based on the red-green bias we calculate for dark pixels in a scene, which divide the datasets into outdoor and indoor environments. Variation of brightness calculation methods did not produce similar patterns for $\alpha_{\%\Delta}$.

We show (by way of proof-of-concept) that an adaptive model based on environmental properties can increase the portability of shadow removal algorithms, by automatically calibrating an algorithm to an environment. The implemented proof-of-concept is performed on the Physical model for shadow removal, a method that employs unsupervised machine-learning to learn the appearance of a shadow pixel over time. We demonstrate that an adaptive model can still increase the efficacy of shadow removal by tuning the parameters, for even an inherently adaptive shadow removal algorithm. The adaptive model is flexible enough to continuously adapt an algorithm over long periods of time, providing a scalable solution without relying on hard-coding thresholds and parameters.

5.1 Future Work

5.1.1 Modeling Low-contrast SIFT Keypoints

Given the correlation sensitivity shown with regard to $SIFT_{fg/bg}$, it is apparent that a more sophisticated model is required for the representation of low-contrast SIFT features in an image. We believe there are improve-

ments to be made to properly model the low-contrast structural changes seen in shadow regions. One such improvement would be restricting the observed low-contrast features to those found within identified foreground objects, eliminating false positives, i.e., low-contrast features existing outside of shadow regions (primarily caused by pixel-level noise).

5.1.2 Rapid Illumination Change Compensation

The presented adaptive model overcompensates for rapid illumination change, hampering shadow detection and discrimination. The problematic illumination changes are exaggerated by a background model that has not adapted to properly represent the new scene brightness. Bales et al.'s BigBackground is a process for illumination compensation that utilizes stable background features to locally estimate illumination changes for an entire scene. By utilizing BigBackground, we can compensate for the sluggish background model, and realistically represent the exaggerated attenuation between shadow and background.

5.1.3 Classifying Indoor/Outdoor Scenes

While implementing various brightness calculation methods, we found (for the α_{dB} model of attenuation) that we could assign an ideal brightness model to a scene based on whether or not it was an indoor or outdoor environment. We quantified the red-green bias of possible shadow pixels, and used the

measured bias to classify a scene as outdoor or indoor. Further verification is required for this claim. Repeating these bias calculations on a wide variety of datasets will validate the spectral-illuminant theory supported the supposition. Further work in this area allows us to study the multi-illuminant properties of outdoor shadows vs. indoor shadows in more depth, which can in turn provide more accurate classification.