**THESIS RESEARCH SUMMARY**

Jay Danner

**The Problem:**

Object tracking and segmentation are pivotal to most preeminent computer vision applications, ranging in application from security and surveillance, to traffic monitoring and analysis. Notably, many applications utilize the extraction of foreground pixels to capture moving objects in a scene; 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 alongside the desired object pixels [1]. While this may occasionally be the intention, shadows generally contribute to inaccurate object classifications and impede proper tracking of foreground objects. In order to rectify this, there have been a bevy of developments regarding shadow detection and removal from foreground objects. In order to perform optimally, these leading methods require assumptions to be made about key factors of a scene, including illumination changes, color variance, clearly delineated foreground objects, etc. As a result, no leading shadow removal methods are robust enough to compensate for a scene over time, nor are they suitable for deployment in an environment without a priori tuning of parameters. The objective of this research is to develop a framework capable of understanding salient scene parameters that affect shadow removal, how they change over time, and how they can be compensated for by modifying shadow removal methods and their parameters.

**Origin + History:**

Previous research pertaining to this research topic include research regarding shadow removal itself, threshold selection for shadow removal, scene characteristics, and scene modeling. This section will also explore the little precedent found in both intelligently selecting algorithm parameters based on scene content, and intelligently selecting between multiple shadow removal methods in a hybrid scheme.

Shadow removal is a necessary step in correct segmentation of foreground objects for proper tracking, as shadows are often brought into the foreground of scene processed with traditional background modeling [3, 4]. 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 2 [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 yield better results.” A second algorithm survey conducted by Sanin et al. in [1] evaluated more modern methods (enumerated 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. The proposed thesis research will use the same datasets as the two surveys. Mitra et al. provides a survey of threshold selection strategies for identifying shadows in moving foreground objects [3].

Characterization of a scene or environment is traditionally done with global parameters such as global hue, saturation, and value (HSV), or color variance. However, more scene properties are needed for proper identification of semantic scene content, needed for this research. The ‘gist’ of scene, proposed by Oliva et al. in [5], compares evaluations of the properties of scene, using low-level features and simple arrangement of volumetric forms. Similarly, in [6], Oliva et al. evaluates the content of a scene using shape modeling such as the openness, ruggedness, roughness, expansion of a scene, creating the ‘spatial envelope’ of a scene. Lowe et al. propose the scale-invariant feature transform (SIFT), which utilizes and identifies features of a scene that are robust to noise, illumination, distortion, and viewpoint [7]. Bayesian Scene Modeling, proposed by Fei-Fei et al., utilizes the ‘bag-of-words’ strategy, building a dictionary of codewords relevant to a scene via unsupervised machine learning [8]. This approach creates a ‘theme’ for a scene built upon these collected codewords, categorizing complex diverse scenes into 13 categories (highway, coast, mountain, etc).

There is precedent for intelligently adapting parameters for shadow removal. Sanin et al., in the implementation of Large Region Texture shadow removal, measure the average global attenuation, saturation, and foreground object perimeter of a scene. The algorithm selects several differing parameter values for hue, saturation, and value thresholds based on the average shadow attenuation and saturation of the scene. It also modifies the texture differential radius needed to qualify a pixel as shadow based on the perimeter of the foreground object in question.

**Proposed Research:**

No leading shadow removal methods are robust enough to compensate for a scene over time, nor are they suitable for deployment in an environment without a priori tuning of parameters. The research proposed will solve this problem by establishing an understanding of scene parameters that negatively affect shadow removal using leading methodology, and use that understanding to perform optimal shadow removal in an arbitrary environment. This requires the creation of a hybrid scheme consisting of two stages: a macro stage which selects the proper shadow removal method apropos of ambient scene properties, and a micro stage that automatically tunes the selected method to optimally perform given the collected scene properties.

The prior assumptions that are required for leading shadow removal methods to perform optimally very rarely lend themselves to quantization or scalability; e.g., a geometric method exercised in this research performs optimally in situations with clearly defined upright objects with predictable shapes. Methods like this one require more of an understanding of semantic scene content rather than lower level measures such as color saturation. Therefore, the technical challenge of the proposed research lies in properly quantifying salient scene properties and content, especially in how these properties pertain to algorithmic methods such as shadow removal. The algorithmic parameters of the selected shadow removal methods can number more than 30 for a particular method. Accordingly, the proposed research requires relatively large-scale sensitivity testing for these parameters in different scenes. The large number of algorithmic parameters incites the need for identifying not only how these parameters affect shadow removal according to the scene's environment, but also how these parameters affect each other.

Lastly, in order to create a framework for shadow removal divorced of previous assumptions, the framework requires an understanding of how soon-to-be identified scene parameters change over time. There is a need for adaptivity in most applications requiring shadow removal, and in the process of obtaining the diverse data sets necessary for this research, the degree to which certain scene parameters change over time can be observed.

**Work Completed:**

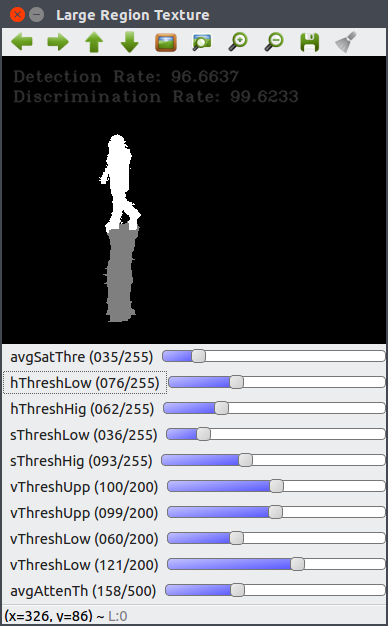
Much of the work completed up to this point has been an empirical study, which has involved development of tools and processes for analyzing, comparing, and evaluating shadow removal methods in differing environments. Standardized implementations of popular shadow removal methods, including ground truths, backgrounds, and frames, are used courtesy of A. Sanin, C. Sanderson, B.C. Lovell (http://arma.sourceforge.net/shadows), licensed under GPL v3+ and written in C++. These shadow removal methods are as follows:

* Chromacity: an approach contingent on color constancy and brightness variance.
* Physical: an approach centered around attenuation based on the color of the source illumination (white for indoors, blue/white for outdoors, etc).
* Geometry: an approach that evaluates the moments of mass of foreground objects, and attempts to find the shadow pixel based on projective geometry.
* Small Region Texture: an approach that attempts to match shadow pixels based on the underlying background texture. This approach achieves this using Gabor filtering.
* Large Region Texture: a similar approach to that of Small Region, Large Region matches shadow pixels based on direction gradient matching.

A more complete taxonomy of these approaches will be provided with the research. The datasets used by Sanin et al. were also provided, and as such will be utilized in this research for more direct and accurate evaluations. The datasets vary in terms of environment, lighting, time of day, and actors of the scene. These different qualities lend themselves to be quite distinct from one another in terms of shadow length, color, orientation, and definition. Four datasets are indoors, and three outdoors. The spatial environments also differ in key factors such as background textures and color variances.

The implementations of the various approaches came with hardcoded parameters that were found to be correspondent with the provided datasets. Prominent parameters include:

* Hue, Saturation, and Value thresholds for color evaluation.
* Global Attenuation thresholds.
* Average foreground object sizes.
* Gradient direction thresholds.
* Canny edge thresholds.

There are many more mutable parameters that affect shadow removal efficacy, and again, a more complete taxonomy will be provided. In order to then quickly evaluate the effect of modifying these in-built parameters, a graphical interface was created to adjust them during runtime and view the effects. The interface supports either entire sequences (e.g. video) or singular frames. The interface supports any of the aforementioned shadow removal methods.

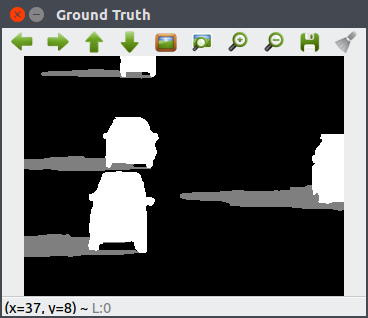
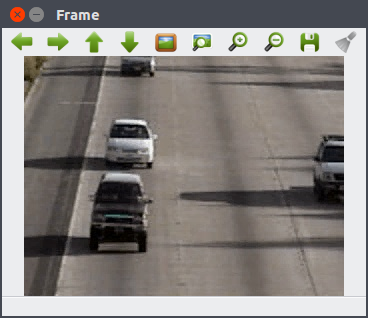
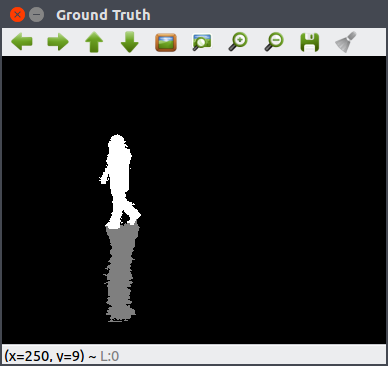
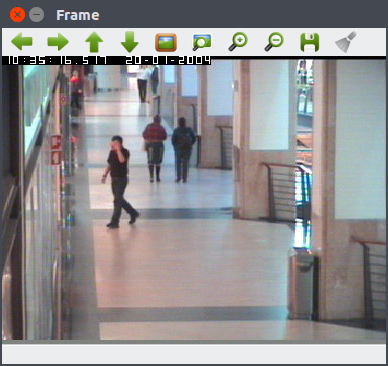
*Fig. 1. Graphical interface for permuting parameters.*

The efficacy of any method on any frame is then evaluated with the popular metrics Detection and Discrimination (seen below). These formulae 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 represent the TP or FN is in reference to (s)hadow pixels of (f)oreground pixels.

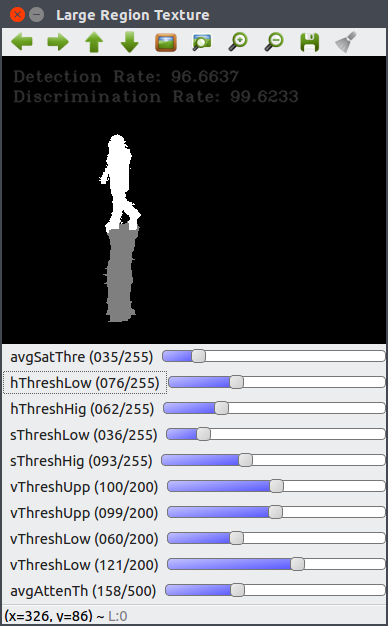
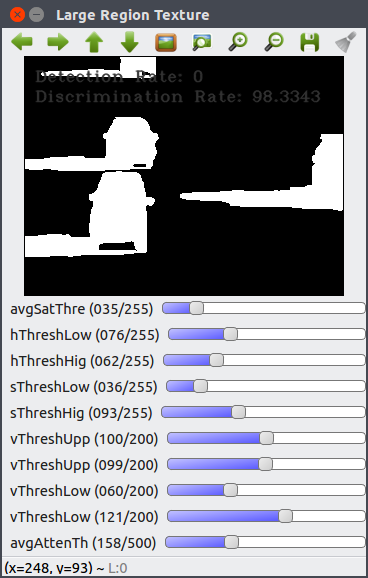
*Detection Rate: η =*

*Discrimination Rate: ξ =*

The graphical interface is a crucial development in terms precisely measuring each parameter's effect on a given scene. However, in order to study shadow removal in terms of adaptivity over time, a more procedural method was required. Courtesy of username ‘brofield’, SimpleINI (github.com/brofield/simpleini), licensed under MIT, allowed for creation of .ini files containing any given removal method's default parameters. The algorithms themselves were then modified to allow for runtime modifications to be made to these parameters, meaning a procedural, iterative approach to analysis was created. A python script capable of writing to this .ini file allows for rapid permutation of parameters and values.

Initial investigation has begun, using the created infrastructure. The thrust of the initial research was to first and foremost prove the problem claimed above, that these leading algorithms are not suitable for arbitrary deployment, and they can be improved with parameters corresponding to their environment. As an example, a parameter dictating the lowest allowable value (HSV) for determined shadow regions is modified.

*Fig. 2. Frames and ground truths.*

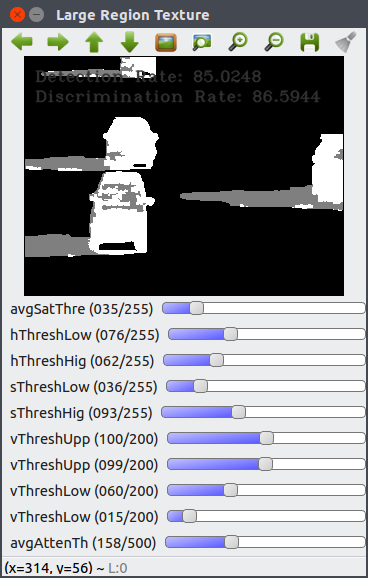
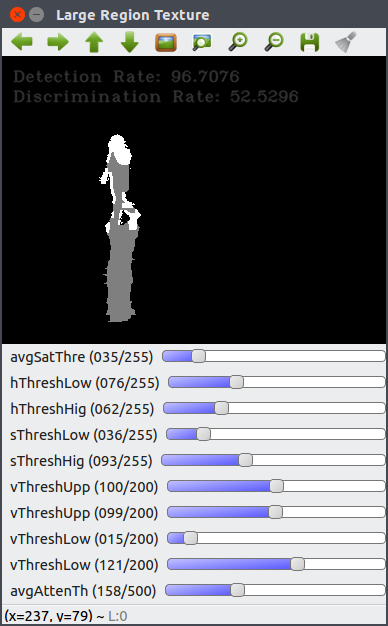
For this demonstration, gray pixel represent detected shadows, with white pixels representing foreground object pixels. For quantitative evaluation, the Detection and Discrimination rates are included in the frame. The demonstrated parameter (*vThreshLower*) causes the LR Texture removal method to perform optimally in the CAVIAR scenes at its default value, 121. Similarly, the same algorithm performs poorly in the included dataset aton\_highway1 with the default parameters.

*Fig. 3. Default parameters.*

*CAVIAR Detection: 96.6637, Discrimination: 99.6233*

*aton\_highway1 Detection: 0, Discrimination: 98.3343*

However, if *vThreshLower* is modified from 121 to 15, the following is observed:

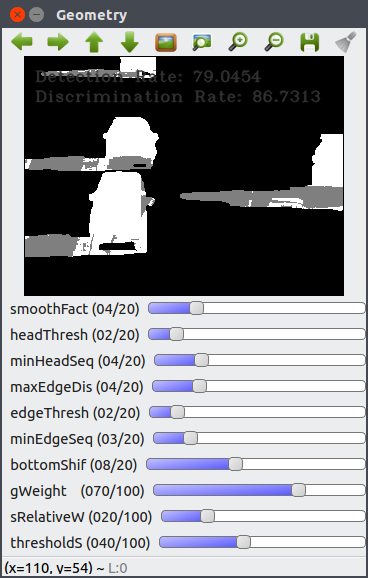
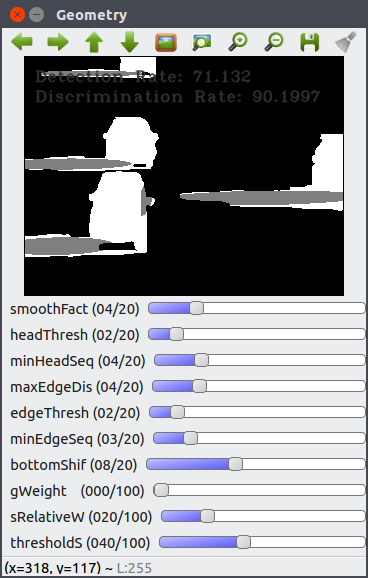


*Fig. 4. Parameter set to ’15.’*

*CAVIAR Detection: 96.7076, Discrimination: 52.5296*

*aton\_highway1 Detection: 85.0248, Discrimination: 86.5944*

As evident above, the parameter change causes the same algorithm to perform significantly better in the aton\_highway1 frame, and poorly in the CAVIAR frame. This justifies the need for the 'micro' stage of the proposed research, in which the algorithms' parameters would be adequately tuned according to their environment. For a second demonstration, the Geometry shadow removal method was found to also showcase different results on the same scene, but with differing parameters.



*Fig. 4. ‘gWeight’ modification.*

*gWeight=70 Detection: 79.0454, Discrimination: 86.7313*

*gWeight=0 Detection: 71.132, Discrimination: 90.1997*

By changing such parameters as the saturation thresholds, changes in the shadow removal efficacy is evident by both increases and decreases in accuracy. Each shadow removal method has been found to have hard-coded, mutable parameters that affect the shadow removal in a given environment.

**Work Remaining:**

The initial thrust of research has been to identify instances of the stated problem. Several high-level qualitative goals need to be satisfied before the creation of a product or conclusion regarding adaptive shadow removal; specifically, a full understanding of algorithm parameters and their effects, and a full understanding of scene parameters are both needed in order to build a correlation between the two strong enough to support an adaptive mechanism. The steps required are broken out explicitly below.

Develop comprehensive understanding of shadow removal method sensitivity to certain algorithmic parameters (e.g., Canny edge thresholds).Before real correlations and conclusions can be drawn, an encompassing understanding of how each algorithm functions is required. This will be done both by simple observation as well as with large-scale batch testing (modifying .ini files containing parameters at runtime).

Develop holistic scene content understanding in order to select shadow removal method for the ‘macro’ stage of hypervisor program. A quantization of seemingly qualitative scene content is required to better select between shadow removal methods. Some of this scene content includes shadow orientation vs. foreground object orientation, pre-existing static scene shadows, foreground object classification, distinctiveness of foreground objects (color differences in particular), light source in a scene, and many more. Building up an understanding of these scene properties may also help with the ‘micro’ stage of the proposed framework.

Develop understanding of how these scene/environment parameters change over time of day and between locales. Observation of how both the quantitative and qualitative scene parameters change over time is required to create the proposed adaptive framework. This step will be simple for things like shadow intensity and global saturation, but more difficult for aforementioned properties like static shadows and object classification.

Create correlation between global scene/environment properties and shadow removal accuracy. This step represents the main experimental thrust of the proposed research detailed thoroughly in this research summary.

Use the aforementioned correlation to build hypervisor program to deploy adaptive shadow removal in arbitrary scenes and environments. This will be done primarily in C++, using the OpenCV API.

Collect appropriate long-term data for adaptivity research. There is a lack of long-term datasets for evaluation and experimentation within this research field; therefore, long-term video sequences must be collected for development. The sequences must reflect significantly shifts in time-of-day, as well as environments. Most of the collected data will be outdoors, although for development and research purposes indoor datasets are utilized.

**Facilities Needed:**

Any facilities needed for the proposed research all correspond to the collection of long-term video sequences, as detailed in the previous section. Therefore, there is a need for recording equipment (camera and tripod), and permission for recording outdoor sequences. If the data recorded must be restricted to campus, this will not significantly impact the research development.

Bibliography

[1] Sanin, Andres, Conrad Sanderson, and Brian C. Lovell. "Shadow detection: A survey and comparative evaluation of recent methods." Pattern recognition45.4 (2012): 1684-1695.

[2] Prati, Andrea, et al. "Detecting moving shadows: algorithms and evaluation."*IEEE transactions on pattern analysis and machine intelligence* 25.7 (2003): 918-923.

[3] Mitra, Bhargav Kumar, Rupert Young, and Chris Chatwin. "On shadow elimination after moving region segmentation based on different threshold selection strategies." *Optics and Lasers in Engineering* 45.11 (2007): 1088-1093.

[4] Nadimi, Sohail, and Bir Bhanu. "Physical models for moving shadow and object detection in video." *IEEE transactions on pattern analysis and machine intelligence* 26.8 (2004): 1079-1087.

[5] Oliva, Aude. "Gist of the scene." *Neurobiology of attention* 696.64 (2005): 251-258.

[6] Oliva, Aude, and Antonio Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." *International journal of computer vision* 42.3 (2001): 145-175.

[7] Lowe, David G. "Distinctive image features from scale-invariant keypoints."*International journal of computer vision* 60.2 (2004): 91-110.

[8] Fei-Fei, Li, and Pietro Perona. "A bayesian hierarchical model for learning natural scene categories." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. IEEE, 2005.