THESIS

AUTOMATIC PREDICTION OF INTEREST POINT STABILITY

Submitted by

H. Thomson Comer

Department of Computer Science

In partial fulfillment of the requirements

for the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2009

COLORADO STATE UNIVERSITY

April 7, 2009

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY H. THOMSON COMER ENTITLED AUTOMATIC PREDICTION OF INTEREST POINT STABILITY BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

Committee on Graduate Work

Dr. Patrick Monnier

Dr. Ross Beveridge

Advisor Dr. Bruce Draper

Department Chair Dr. Darrell Whitley

ABSTRACT OF THESIS

AUTOMATIC PREDICTION OF INTEREST POINT STABILITY

Many computer vision applications depend on interest point detectors as a primary means of dimensionality reduction. While many experiments have been done measuring the repeatability of selective attention algorithms [MTS+05, BL02, CJ02, MP07, SMBI98], we are not aware of any method for predicting the repeatability of an individual interest point at runtime. In this work, we attempt to predict the individual repeatability of a set of 10^6 interest points produced by Lowe's SIFT algorithm [Low03], Mikolajczyk's Harris-Affine [Mik02], and Mikolajczyk and Schmid's Hessian-Affine [MS04]. These algorithms were chosen because of their performance and popularity. 17 relevant attributes are recorded at each interest point, including eigenvalues of the second moment matrix, Hessian matrix, and Laplacian-of-Gaussian score.

A generalized linear model is used to predict the repeatability of interest points from their attributes. The relationship between interest point attributes proves to be weak, however the repeatability of an individual interest point can to some extent be influenced by attributes. A 4% improvement of mean interest point repeatability is acquired through two related methods: the addition of five new thresholding decisions and through selecting the N best interest points as predicted by a GLM of the logarithm of all 17 interest points. A similar GLM with a smaller set of author-selected attributes has comparable performance.

This research finds that improving interest point repeatability remains a hard problem, with an improvement of over 4% unlikely using the current methods for interest point detection. The lack of clear relationships between interest point attributes and repeatability indicates that there is a hole in selective attention research that may be attributable to scale space implementation.

H. Thomson Comer
Department of Computer Science
Colorado State University
Fort Collins, CO 80523
Spring 2009

TABLE OF CONTENTS

1 Introduction **1**

1.1 Interest points in computer vision 1

1.2 Measuring interest points 3

1.3 Interest point repeatability prediction 4

2 Literature Review **6**

2.1 Selective attention as signal theory 6

2.1.1 Scale spaces 7

2.1.2 Scale invariance 7

2.1.3 Affine invariance 9

2.2 Selective attention and biology 9

2.3 Evaluation and review 11

3 Implementation **13**

3.1 Use of scale invariant algorithms 14

3.2 Scale-space 14

3.3 DOG 17

3.4 Harris-Laplace 18

3.5 Hessian-Laplace 19

3.6 Interest point comparison metrics 20

3.6.1 Repeatability 20

3.6.2 Accuracy 21

3.7 Implementation differences and discussion 21

4 Experiments **23**

4.1 Attributes of interest points 24

4.2 Attribute thresholding 26

4.3 Logistic regression 28

4.3.1 Normalization techniques 30

4.3.2 Regression by interest point detector 31

4.4 Individual attribute performance 35

4.4.1 Scale 37

4.4.2 Harris eigenvalues 38

4.4.3 Hessian eigenvalues 43

4.4.4 Entropy scores 47

4.4.5 Values of extrema and their neighborhood 50

4.5 Extrema inversion 54

4.6 Method of extrema detection 55

5 Conclusion **58**

5.1 Summary experiments 59

5.1.1 Thresholds 59

5.1.2 Multivariate generalized linear modeling 60

5.2 Discussion 64

5.3 Future work 65

References **67**

LIST OF FIGURES

1.1 A small set of highly relevant interest points suitable for face recognition. . 3

3.1 Two views of an image pyramid. 16

3.2 An extra level in each octave of a scale space is used to produce a Difference-of-Gaussians Pyramid. 18

4.1 Repeatability of interest points thresholded by the Hessian determinant as suggested by Lowe [Low03]. Repeatability is maximized by discarding interest points with a negative Hessian determinant. 26

4.2 Relationship of repeatability and ratio of Hessian eigenvalues. Repeatability is maximized for interest points where $r \leq 5$ regardless of the algorithm used for detection. 27

4.3 Repeatability of interest points with Harris R values above a threshold. We did not find a Harris threshold that improves repeatability. These results show that performance decreases as R increases. We also examined the accuracy of Harris interest points with a similar result. 28

4.4 Logistic regression by normalization type. The area under the curve (AUC) for each attribute on original data and each of three common normalization techniques. Fitting determinant of Harris $hardeterminant$ and optimized value $truevalue$ attributes fail because of the magnitude of these attributes. 32

4.5 Logistic regression by normalization type. The correlation $r_{E(Y),Y}$ for each attribute on original data and each of three common normalization techniques. Log normalization reduces quality of fit for only two attributes and causes $r_{E(Y),Y}$ and AUC to correspond highly ($p < 0.0000001$). . . 32

4.6 Regression by algorithm on original attributes. Performance of the GLM predictions are low but non-random. The fitting of Hessian interest points maximizes $r_{E(Y),Y}$ and the Harris fitting maximizes AUC. 34

4.7 Regression by algorithm on log attributes. Log normalization introduces uniformity and indicates DOG is most predictable. 34

4.8 Extremum near the borders of images are predictably not as repeatable. $r_{E(Y),Y} = 0.01, AUC = 0.51$ 35

4.9 Logit function predicted by the GLM, ROC curve, and conditional density estimation of scale. Hessian points are the most stable to scale increases, with the most stable points at the bottom of an octave and the least stable at the top. $r_{E(Y),Y} = 0.03, AUC = 0.52$ 37

4.10 Investigation of repeatability of interest points against the ratio of Harris eigenvalues. Our results show repeatability of almost 90% for interest points with eigenvalue ratio below 5. 39

4.11 Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $harlambda1$. Repeatability decreases as the first eigenvalue increases as interest points become more like edges and less like corners. High correlation and low AUC suggest a bad fit: $r_{E(Y),Y} = 0.07, AUC = 0.51$ 40

4.12 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $harlambda1$. $r_{E(Y),Y} = 0.00, AUC = 0.49$ 40

4.13 Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $harlambda2$. High AUC and low correlation suggest overfitting of the model: $r_{E(Y),Y} = 0.02, AUC = 0.55$ 41

4.14 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $harlambda2$. Collapsing the variance reveals a clear linear relationship for all three algorithms. The most predictive attribute: $r_{E(Y),Y} = 0.08, AUC = 0.55$ 41

4.15 Logit function predicted by the GLM, ROC curve, and conditional density estimation of $hardeterminant$. The large difference in slope for DOG is caused by variance, seen in the next figure. $r_{E(Y),Y} = 0.0018, AUC = 0.48$ 42

4.16 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $hardeterminant$. $r_{E(Y),Y} = 0.04, AUC = 0.52$ . . 42

4.17 Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $heslambda1$. $r_{E(Y),Y} = 0.02, AUC = 0.54$ . . . 44

4.18 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $heslambda1$. $r_{E(Y),Y} = 0.07, AUC = 0.55$ 44

4.19 Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $heslambda2$. This attribute increases linearly with repeatability and suggests discarding when < 0. $r_{E(Y),Y} = 0.07, AUC = 0.55$ 45

4.20 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of Hessian $heslambda2$. The strong relationship from the original feature disappears after the absolute value is taken in log normalization. $r_{E(Y),Y} = 0.03, AUC = 0.51$ 45

4.21 Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $hesdeterminant$. The slopes are exaggerated because of high variance and are reduced in the nexture figure. $r_{E(Y),Y} = 0.01, AUC = 0.54$ 46

4.22 Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $hesdeterminant$. $r_{E(Y),Y} = 0.05, AUC = 0.53$ . . 46

4.23 Logit function predicted by the GLM, ROC curve, and conditional density estimation of entropy. Interest points with $entropy < 1$ should be discarded. $r_{E(Y),Y} = 0.03, AUC = 0.52$ 48

4.24 Logit function predicted by the GLM, ROC curve, and conditional density estimation of first derivative of entropy. Interest points with $dentropy > -1$ are 4% less repeatable than others. $r_{E(Y),Y} = 0.05, AUC = 0.54$ . . 49

4.25 Logit function predicted by the GLM, ROC curve, and conditional density estimation of second derivative of entropy. $r_{E(Y),Y} = 0.02, AUC = 0.52$ 49

4.26 Logit function predicted by the GLM, ROC curve, and conditional density estimation of $value$ at each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.03, AUC = 0.52$ 52

4.27 Logit function predicted by the GLM, ROC curve, and conditional density estimation of sub-pixel optimized $truevalue$ at each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$ 52

4.28 Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$ 53

4.29 Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$ 53

4.30 Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.00, AUC = 0.51$ 54

4.31 Two types of extrema detection. Level extrema are detected in the Harris or Hessian signal and tower extrema are detected in the Laplacian-of-Gaussian in [MS02]. Extrema are detected more rigorously in [Low03] using cube extrema. Use of cube extrema greatly reduces the number detected and negatively affects the repeatability of H-L interest points. . 56

5.1 Fitting of a GLM fit to six author selected attributes and to a GLM fit to all 17 attributes including log normalized Harris, Hessian, and value families. 63

5.2 Fitting of a GLM fit to five author selected attributes and to a GLM fit to all 17 attributes including log normalized Harris, Hessian, and value families. 63

LIST OF TABLES

4.1 Initial results verifying expected repeatability rates and interest point density of each algorithm. 23

4.2 Extrema inversion results. Negatively valued extrema are slightly less repeatable than positive extrema [Low03]. 55

4.3 Initial results with cube neighborhood extrema detection constraint for H-L algorithms. These data are produced from a different set of randomly selected source images. The ratio of interest point density is informative. 56

5.1 Log odds coefficients produced by a GLM trained to predict the repeatability of an interest point. 62

Chapter 1

Introduction

Computer vision applications typically depend on extracting information from large, complex visual scenes. One popular approach is to reduce the data by concentrating on a small set of interest points (also referred to as selective attention windows, local features, image regions, keypoints or extrema). Interest points are distorted less by changes in viewpoint than the full image and therefore provide repeatable clues to the contents of the scene. In this work we model the attributes of interest points in order to predict the accuracy and repeatability of individual interest points at runtime.

1.1 Interest points in computer vision

Interest points are a successful approach to reducing the dimensionality of a scene in computer vision applications. David Lowe's SIFT [Low03] and Mikolajczyk and Schmid's Harris and Hessian-Laplace [MS02] algorithms have been cited cumulatively 4580 times since their release according to Google Scholar. These algorithms use an efficient gradient-based technique to produce a set of interest points - informative image sub-regions that are localized in scale and space.

Interest points are popular for a number of reasons. They are efficient, producing a representative sampling of an image with near realtime performance. They are robust to transformation and viewpoint change because they depend only on local struc-

ture. Interest points can also be produced independently of segmentation techniques or any discussion of foreground/background separation. Computer models of human vision utilize interest points because of their similarity to psychological theories of vision [Dun98, GSKE+99, IKN+98, KU85, MP90, OFPK02, OvWHM04, PLN02, PIIK05]. Receptive-field models [dB93, JP87, SH85] suggest that categorization and classification in human cortical areas correspond to specific viewpoint locations and signal structures. Numerous models of human vision using interest points and their corresponding research basis are discussed in Chapter 2.

An algorithm is considered an interest point detector if it depends on a set of characteristics shared between numerous approaches. Interest point detectors produce local image sub-regions located in scale and space and focused around corners, blobs, and image curvature. They provide a broad sampling of content in the original image while being independent of global scene information. Interest points can be produced RANSAC-style with high density and low repeatability, or in lower density by focusing on the points with the strongest attributes. Descriptors are often used in conjunction with interest points in order to model local image regions for object recognition [ECV, Low99, OPFA06], segmentation [JLK], or face recognition [BLGT06, MBO07, KS] as in Figure 1.1. Global scene content can also be modeled with interest points using a combination of descriptor information and the spatial relationships between them [SC00, SM97, SL04]. Scenes can be reconstructed using the interest points as a set of individualized scene anchors containing global spatial and local image content information [SI07].

Interest points provide local access to global information structure. They do this using fast, deterministic algorithms that are able to reliably select the same set of interest points with high repeatability from a variety of image scenes. This and other good attributes of interest points make them useful in a wide variety of image processing

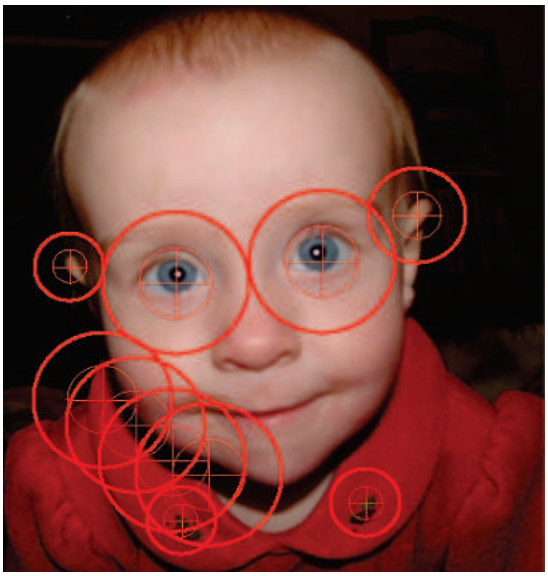

Figure 1.1: A small set of highly relevant interest points suitable for face recognition.

applications.

1.2 Measuring interest points

Interest points have a number of properties that make them useful for computer vision tasks. Tuytelaars and Mikolajczyk provide a detailed description of these properties in [TM08], some of which are repeated here.

- Repeatability: A repeatable interest point appears on the same structure in a pair of images taken from different viewpoints. This is the most common metric in interest point stability evaluation.
- Accuracy: Detected points are accurately localized in scale, shape, and space. Registration problems use only the interest points with highest accuracy for calibrating epipolar geometry.
- Locality: Interest points are localized in scale and space using local structure information that is robust to scene change.

- Distinctiveness: The degree to which an interest point represents local image structure. This interest point characteristic is often tightly coupled with an interest point descriptor.

- Density: Interest points should be produced in such a quantity that informative areas of image structure are well sampled. In addition, algorithms that produce more interest points will produce more informative subsets of those points.

- Efficiency: Interest points are often used in real-time computer vision systems and as such the algorithms for their detection should be fast.

An interest point can be evaluated on its information content as well as its repeatability. The information content of an interest point can not be easily measured without ground-truth specifications. The repeatability of an interest point, however, can be measured easily. This is the measure used to evaluate interest points in this work.

1.3 Interest point repeatability prediction

We seek to predict the repeatability of interest points using attributes from three pivotal algorithms in selective attention research. The Difference-of-Gaussian (DOG) interest point detector from Lowe's SIFT algorithm is the single most commonly used algorithm. It has been cited 2.4 times per day since its publication. It selects interest points that are the extrema in the DOG filter response computed over an image pyramid. Mikolajczyk and Schmid [Mik02, MS02] propose two highly-repeatable algorithms as the basis of their affine-invariant technique. The first of these, the Harris-Laplace, detects interest points using the Harris corner measure, computed from the second moment matrix of first derivatives, and the Laplacian-of-Gaussian (LOG) filter, an analog of the DOG. The Hessian-Laplace detects interest points using a combination of the Hessian matrix of second derivatives and the LOG and is the best performing derivative-based technique

found in a comparison of affine-invariant detectors [MTS+05]. We are not interested in performing another black-box comparison of interest point algorithms; for such comparisons, see [DL04, MTS+05, MLS05]. Instead, we are interested in measuring attributes of individual interest points (no matter what algorithm they were generated by) and in determining whether the repeatability of an interest point can be predicted based on these measures.

Each selective attention algorithm lacks a means for predicting the usefulness of individual interest points at run-time. The algorithms produce many interest points; in most applications users keep and process only a subset of them. Interest point repeatability prediction will therefore provide users with the ability to parameterize each algorithm towards either increased quantity or increased repeatability. It will also lead research techniques to generate more repeatable algorithms by focusing on the attributes that improve individual repeatability most.

This work measures the overall and algorithm-specific repeatability of one million interest points produced by the above three algorithms from randomly selected images from the CalTech-101 database [FFFP07]. This analysis has several goals. One is to determine which attributes, if any, are predictive of whether an interest point will be detected again in another image, and whether the current thresholds in common use are well chosen. Another goal is to determine to what extent the repeatability of an interest point can be predicted from its bottom-up properties. Finally, our third goal is to fit a multi-attribute statistical model to best predict which interest points will repeat.

Chapter 2

Literature Review

Many selective attention algorithms have been proposed. The algorithms take inspiration from separate schools of psychology and psychophysics, signal theory, and information theory. Many of the following works combine the efforts of multiple schools, producing one of the most interesting topics in computer vision. There are two primary schools responsible for selective attention research: those of signal theory and those of biology.

2.1 Selective attention as signal theory

Original work produced interest point detectors like corner detectors, edge detectors, the computation of principal curvature in an image, and information points based in information theory. These algorithms were extended to multiple scales using the concept of scale invariance and the computation of a scale space. Scale space algorithms detect interest points at multiple parameterizable levels of scale and radius. Recent work has extended scale space algorithms further into affine invariance. These algorithms find ellipsoidal instead of circular interest points at multiple levels of scale.

2.1.1 Scale spaces

Interest points detected on the original image signal only sample its smallest frequency range. Many interest point detectors attempt to extract scale-invariant interest points through use of a "scale space". Seminal work by Koenderink [Koe84] and Witkin [Wit87] provide the basic framework for such scale invariant interest points. In their work scale spaces are computed from a set of derivative-normalized Gaussian convolutions of a source image. Using a scale space representation, interest points can be detected at any required scale. The scale space of an image can be computed over the original image, derivatives of the original image, or by calculating the entropy at each pixel coordinate in x, y and scale. Eaton et al. give detailed instructions for building a scale space used with scale-invariant interest point generators [ESM+06] and Burt and Adelson discuss their theoretical foundations [BA83].

2.1.2 Scale invariance

Lindeberg [Lin94] suggests methods for locating the characteristic scale in an image scale space by producing an image pyramid from successive convolutions with a gaussian. Extrema of the "Laplacian-of-Gaussian" (LOG) function across levels of the scale space find optimal locations in the second order derivative of the image in both x, y coordinates as well as region size. An optimal region, or characteristic scale, also gives a local frequency estimation. This work is extended [Lin98] by finding and annotating the characteristic scale of local image structures including blobs, junctions, and ridges.

David Lowe contributes to the understanding of Lindeberg's previous work by showing that the LOG operator can be approximated with a Difference-of-Gaussians (DOG) pyramid [Low03]. Lowe's DOG pyramid is able to find extrema in the scale space of an image similar to those found by Lindeberg's LOG. The run time of Lowe's DOG function is significantly improved over LOG by eliminating the convolution with an

LOG filter. The DOG approach is the keypoint detection stage of an algorithm called Scale Invariant Feature Transform (SIFT), a popular constellation-of-features based object recognition technique. A reimplementation of the DOG interest point detector is one of the algorithms analyzed in this work.

Lowe's SIFT algorithm is used often in subsequent publications, including Ledwich and Williams who use SIFT features for image retrieval and outdoor localization [LW04]. Clusters of SIFT local features are used in a Hough space to perform object recognition and perform an 8-dof homography between images. The usefulness of performing sub-pixel optimization via $3D$ quadratic is also demonstrated [BL02]. Other uses of SIFT are numerous and exist for object recognition, face detection, scene reconstruction, and many other applications [FPZ03, OPFA06, WRKP04].

Mikolajczyk proposed a new scale invariant interest point generator in his Ph.D. thesis [Mik02]. The Harris-Laplace detector combines scale-sensitive Harris corners with Lindeberg's detection of the characteristic scale. Harris-Laplace first uses the Harris corner detector to find maxima in the second order moment matrix of first derivatives [HS88]. Those Harris points that are also extrema in the LOG are then accepted as keypoints. He and Cordelia Schmid propose a similar algorithm using the blob detection of the Hessian in place of Harris corner points [MS04]. Our work also examines the features and performance of these two algorithms due to their performance and popularity.

Information-theoretic approaches using Shannon entropy have been used for a variety of image processing applications. Gilles' Ph.D. work applied regional measurement of entropy to aerial images[Gil98]. In order to select scale-invariant interest points, this work was extended by Kadir and Brady to a multi-scale representation[KB01], called Scale-saliency. The use of entropy to detect regions of interest in an image is intuitive since the goal of all selective attention algorithms is to detect the most informative set of regions. The algorithm proposed by Kadir and Brady, called Scale-saliency, finds

regions in an image where the second derivative of entropy with regard to scale is zero. The level of scale where the second derivative of entropy is zero then defines a bounded circular region inside of which the entropy is greater or less than its immediate neighborhood.

Scale-saliency has been used by a variety of authors. Hare and Lewis use the scale-saliency approach for tracking and identifying objects through image matching sequences[HL03], providing $3D$ motion tracking in real time. Fei Fei et al. use Scale-saliency local features to perform constellation-of-features style object recognition [FFFP07], and Fergus et al. use them for object class recognition [FPZ03].

2.1.3 Affine invariance

Mikolajcyk and Schmid proposed another successful selective attention algorithm called Harris-Affine in [MS04]. They extend Harris-Laplace with an iterative algorithm that adaptively fits the keypoints with increasing precision and then fits them with a second moment matrix that defines the bounding ellipse of the extremal region.

Kadir and Brady [KZB04] extend Scale-saliency interest points to affine invariance using an iterative approach over the original scale invariant Scale-saliency points. It is shown to perform similarly to curvature-based techniques with improved performance for small perturbations.

Maximally Stable Extremal Regions (MSER) are interest points generated using a fast watershed algorithm. It has performance comparable to the best affine invariant approaches [MCUP04].

2.2 Selective attention and biology

The use of rapid non-contextual interest point detectors is well supported in biological literature [KB01]. Biological attention research is based on artificial intelligence (A.I.),

visual recognition tasks, and aspects of the growing biomimetic community that seeks to model already proven systems (those we see in nature). Numerous A.I. systems are using interest point generators to make judgements about image content in order to localize the objects viewed in the scene or the actor's position within it. In order to improve these systems and provide an observational justification for their existence, many researchers are turning toward biomimetic models. Selective attention is the first stage in many of these systems, using the research of psychophysics and psychology to model the interest points used in later cortical areas.

There exists a large body of psychology research demonstrating the validity of selective attention systems in human and animal visual systems. Li et al. [LVKP02] show that the identification and categorization of image scenes occurs in the early stages of the visual system, massively and in-parallel. Malik and Perona [MP90] provide the biological foundation for LOG/DOG techniques by proposing a model of human attention based on the differences of offset Gaussians observed in human V1 receptive fields [SH85]. Multiple sparse local features are supported by Tsunoda et al., who show that complex objects are represented as additive features in inferotemporal cortex [TYNT01].

Koch and Ullman proposed the use of a saliency map [KU85]. Based on neurological studies, they suggest that human attention is a sum of saliency maps tuned to various image features. In order to detect the regions of highest saliency, they propose the use of multi-scale DOG filters followed by a winner take all neural network. By using an image pyramid to provide analysis of scale space, the winner-take-all feedback network finds salient regions of varying scale. A neurally inspired, multi-layer neural network based on selective tuning is proposed by Tsotsos et al. [TCW$^+$95]. Interest points are selected via tuning and a winner-take-all neural network.

Itti's Neuromorphic Vision Toolkit (NVT) [IKN$^+$98] combines massively parallel feature detection [TG80] with the combination of multiple feature maps [KU85] to pro-

duce biologically plausible feature maps. Feature maps are computed for opponent-color and intensity channels, and 8 principle orientations. These maps are combined using a scale space similar to Lowe's DOG [Low03] into a single topographic saliency map. Interest points are ranked according to a winner-take-all neural network with suppression. In a time series, this suppression leads NVT to fixate on each interest point in descending rank. Siagian and Itti [SI07] suggest the evaluation of vision applications for speed, performance, and a measure of their biological-ness. This hypothesis is extended to rapid scene classification using the NVT system from Itti.

Peters et al. [PIIK05] extend the bottom-up salience model of selective attention to include interactions between orientation-tuned cells for clutter reduction and contour facilitation. Their work builds on Parkhurst et al. [PLN02] who demonstrate that human eye-tracking can be partially accounted for using a Difference-of-Gaussians model.

Sun and Fischer [SF03] produce a biologically inspired vision system based on Duncan's Integrated Competition Hypothesis, which suggests that early, pre-cortical regions of the human visual system compete in parallel with tuning and later regions for the selection of salient regions [Dun98]. Sun and Fischer use selective attention to compute the visual salience of objects and groupings of objects at an early stage, combining that with a second region that implements hierarchical selectivity of attentional shifts.

2.3 Evaluation and review

A number of evaluations have been made considering which of these two front ends generates more stable keypoints. Mikolajczyk et al.[MTS$^+$05] evaluate the accuracy of interest point detectors against each other under affine transform and find Hessian-Affine and Maximally Stable Extremal Region (MSER) keypoints to be most stable. Draper and Lionelle[DL04] recently compared the performance of two DOG-filter based techniques. Mikolajczyk et al. test various interest point detectors for their usefulness in

object recognition tasks [MLS05]. Object recognition performance is improved with the use of interest points, particularly using those from Hessian-Laplace and Scale-saliency. Descriptors used as the second stage of selective attention algorithms are compared in [MS05]. Mikolajczyk et al. also compare the invariance of affine-interest point detectors, finding MSER and Hessian-Affine interest points the most effective [MTS+05]. Itti and Koch provide a detailed review and justification of attentional models inspired in psychological studies [IK01]. Several computational architectures and their application to objective evaluation of advertising design are reviewed by Itti [Itt05]. Finally, Tuytelaars and Mikolajczyk undertake a broad survey of the history, progression, and implementation of interest point detectors [TM08].

Chapter 3

Implementation

While there exist exhaustive tests of the comparable performance of various interest point detectors [MTS$^+$05, BL02, CJ02, MP07, SMBI98], no experimental demonstration of the selected interest points has been performed. The purpose of this research is not to compare the performance of three fairly well known interest points detectors, but to determine which attributes of those detectors are the most descriptive, and why.

10^6 interest points are generated randomly from images in the CalTech-101 image dataset using three state-of-the-art algorithms and tested for repeatability. Each interest point is produced from one of three algorithms - Lowe's LOG approximation [Low99], hereafter referred to as DOG, and Mikolajczyk and Schmid's Harris-Laplace [Mik02] and Hessian-Laplace [MS02] algorithms. Harris-Laplace produces interest points at corners, Hessian-Laplace produces interest points at circular blobs, and DOG produces interest points at blobs and edges. Hessian-Laplace and Harris-Laplace algorithms are sometimes referred to in this work as H-L algorithms to denote their similarity, and Lowe's algorithm is denoted as DOG because the behavior of his descriptor is not examined.

These algorithms are representative of the state-of-the-art in interest point detection, with one notable exception. Matas et al.'s MSER algorithm [MCUP04] offers affine-invariance, rapid run-time, and good performance, but is not part of the derivative-based

class of algorithms we evaluate.

3.1 Use of scale invariant algorithms

We test the scale invariance class of algorithms instead of the affine invariance class for a number of reasons. Lowe's DOG is the only mechanism that is biologically tested [IKN+98, OvWHM04] and it is the most commonly used interest point operator in modern literature. Scale invariant algorithms are, in general, faster than the affine invariant methods because no procedural iteration is required. Affine invariance commonly uses iteration on mathematical models to locate the interest point region and introduce isotropy

Scale invariant interest points can be viewed as the set of all isomorphic affine invariant interest points. Finding the homography between two affine interest points is equivalent to finding their shared isomorphy. The most predictive characteristics of scale invariant interest points then are good guides to the predictive characteristics of affine invariant interest points.

Finally, the importance of using similarity versus affine invariant interest point detectors is not yet known. Similarity transformations are the simplest form of planar transformation defined by the perspective transformation. Affine transformations are the middle-ground between these two extremes. In either the affine or perspective transform, only small changes are acceptable. This is because finite sampling effects caused by scaling make repeatability impossible. We focus on scale invariance because it shares the goals of affine invariance with a faster runtime, reduced complexity, and a more complete history.

3.2 Scale-space

In order to represent image structures over all scales a scale space must be constructed [Koe84, Wit87]. The scale space gives a discrete representation of the continuous signals present in an image. Detecting extrema in the scale-space of an image provides the scale of an underlying structure [Lin98].

A scale space is produced by successive convolution of a source image I with a Gaussian kernel. Since the Gaussian kernel is separable we can use a 1D Gaussian given by

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \tag{3.1}$$

producing a series of images $L(\sigma, x) = g(\sigma, x) * I$. Each time the Gaussian kernel $\sigma = 2.0$, I is subsampled by a factor of two, reducing the size of I by four while retaining signal and eliminating noise. Thus, a scale space is a pyramid of images where the width and height of each level decrease by two successively.

The appropriate set of σ values used in the scale space depends on the size of image structures we seek to detect. Using $\sigma = 2.0$ produces a very coarse pyramid and only responds to image structures with scales that are powers of two. Therefore, we divide each octave into an integer number of levels s such that the constant scale difference between levels $k = 2^{1/s}$. Our experiments use $s = 3$ based on the results of Lowe [Low03], who found that three levels per octave maximize repeatability. The pyramid is then composed of octaves each containing three levels with $\sigma = \{2^0, 2^{1/3}, 2^{2/3}\}$. The bottom half of Figure 3.1 shows an image pyramid with $s = 3$. The upper half shows two levels in the same octave, and one level each higher octaves. Additional details of pyramid construction are available from Burt and Adelson [BA83] and Eaton et al. [ESM+06].

When the size of an octave becomes smaller than the convolution mask used to

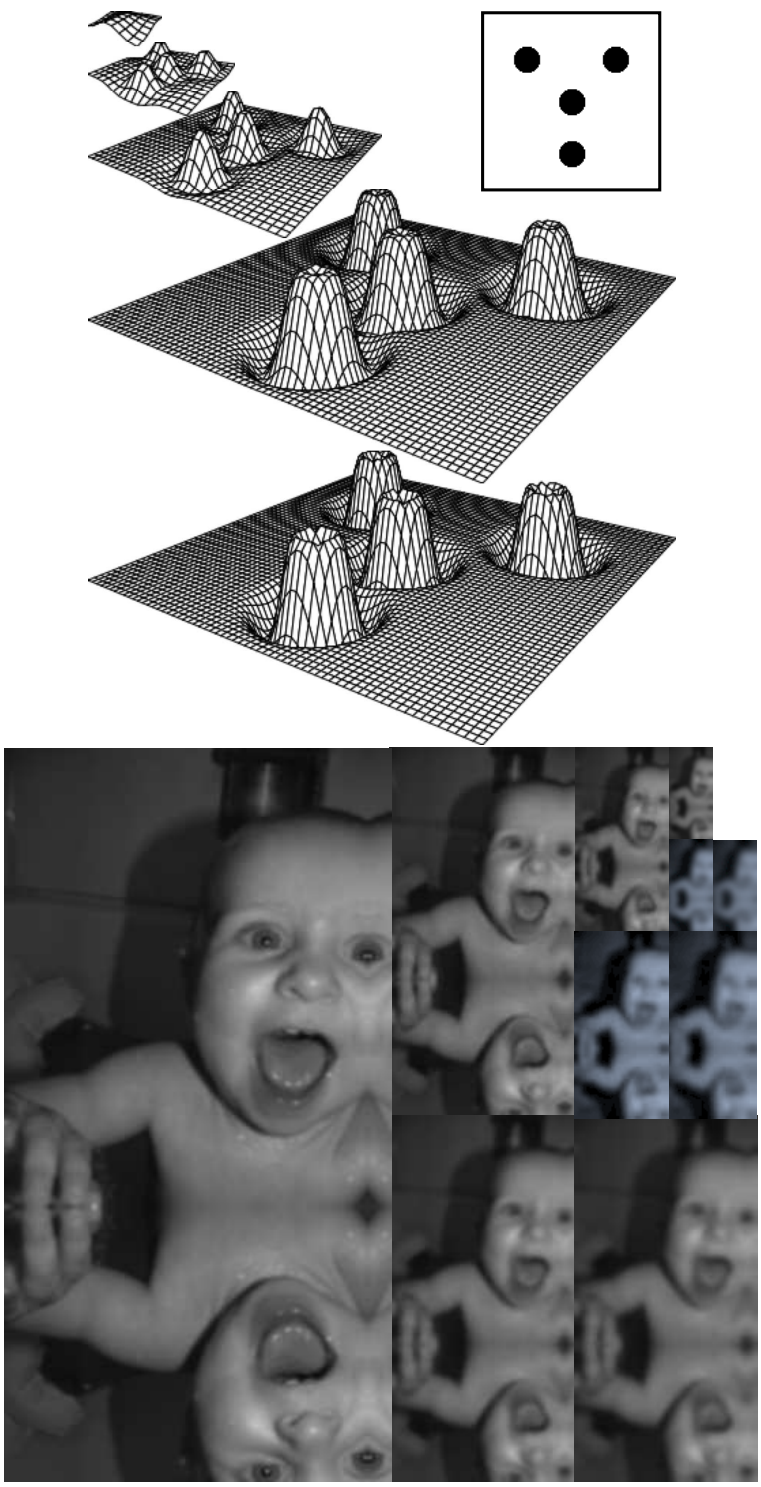

Figure 3.1: Two views of an image pyramid.

produce each successive octave, the pyramid is finished. We use a 9×9 convolution mask at all levels of scale. 9×9 minimizes the error between the expected σ of each level of an octave and the true σ.

3.3 DOG

A set of difference images is produced from the original scale space, producing a DOG-pyramid. This is based off of Lowe[Low99] who demonstrated that the LOG diffusion equation

$$\frac{\delta G}{\delta \sigma} = \sigma \nabla^2 G \tag{3.2}$$

is approximated and optimized for speed with a DOG pyramid. The DOG pyramid improves computation time by eliminating the necessity of derivative convolutions.

The DOG algorithm produces an image pyramid containing a series of difference images D, produced from a pyramid of gaussian images L (Equation 3.4).

$$\begin{aligned} L(x, y, \sigma) &= G(x, y, \sigma) * I(x, y) \\ G(x, y, \sigma) &= \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \end{aligned} \tag{3.3}$$

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{3.4}$$

In order to create a DOG-pyramid, one extra gaussian level (Figure 3.2) is produced, creating $s + 1$ levels per octave. The difference of each set of four levels is taken, producing a difference pyramid with the same size as the traditional scale-space.

By detecting extrema in the difference image D of each pair of images, salient regions are detected in both scale and space. Interest points are located at these extrema.

The scale of an interest point is determined by its position in the image pyramid. Interest points are then localized by fitting a 3D quadratic to the neighboring points around each potential maximum:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x^T}\frac{\partial^2 D}{\partial \mathbf{x^2}}\mathbf{x} \tag{3.5}$$

where $D(\mathbf{x})$ is evaluated at the sample point and $\mathbf{x} = (x, y, \sigma)$ is the offset from this point as explained by Lowe [Low03].

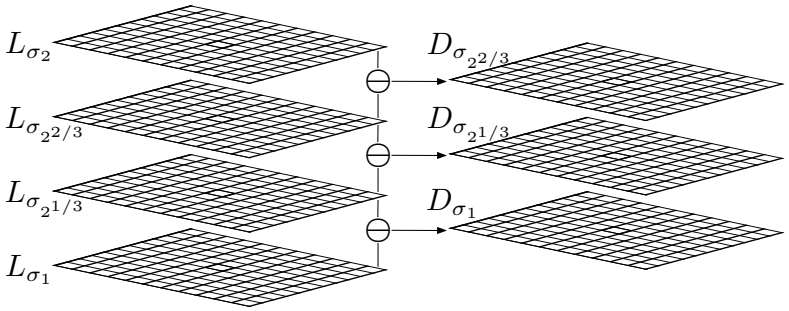

Figure 3.2: An extra level in each octave of a scale space is used to produce a Difference-of-Gaussians Pyramid.

3.4 Harris-Laplace

The Harris corner detector is one of the best and most-tried corner detectors but it has no scale component [HS88]. The Harris-Laplace algorithm uses a scale space to produce interest points by detecting Harris corner points on each level of an image pyramid. The corner points are localized in scale by finding the "characteristic scale" using the Laplacian-of-Gaussian (LOG) [Lin98] filter.

Harris corners are constructed using the second-moment matrix $\mu(x, \sigma_I, \sigma_D)$ in the scale-normalized first derivative of the source image.

$$\mu(x, \sigma_I, \sigma_D) = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix} \tag{3.6}$$

where $\sigma_{I,D}$ are the integration and differentiation kernel sizes, respectively, g is a gaussian, and $L_i^2(x, \sigma_D)$ is the square of the intensity of the first derivative with respect to i at position x.

Interest points are detected using maxima in the Harris measure R

$$R = \det(\mu(x, \sigma_I, \sigma_D)) - \alpha \operatorname{trace}(\mu(x, \sigma_I, \sigma_D))^2 \tag{3.7}$$

The second moment matrix describes the orientation and magnitude of gradients around each candidate interest point. The second moment matrix is the covariance matrix of partial derivatives of image intensity around a candidate interest point and is used for the detection of corners.

3.5 Hessian-Laplace

The Hessian matrix

$$H(\mathbf{x}) = \begin{bmatrix} L(L_x)_x(\mathbf{x}) & L(L_x)_y(\mathbf{x}) \\ L(L_x)_y(\mathbf{x}) & L(L_y)_y(\mathbf{x}) \end{bmatrix} \tag{3.8}$$

is used for the detection of blobs. Instead of the covariance of the first derivative neighborhood, $H(\mathbf{x})$ contains the second derivative information at the exact coordinates of the extremum in $x = I(x, y, \sigma)$, denoted by $L(L_i)_j(x)$. Interest points produced by the Hessian-Laplace detector are simultaneously maximal in the trace and determinant of the Hessian matrix.

$$\mathrm{DET}(H) = \sigma_I^2(L(L_x)_x L(L_y)_y(\mathbf{x}) - L(L_x)_y^2(\mathbf{x})) \tag{3.9}$$

$$\mathrm{TR}(H) = \sigma_I(L(L_x)_x(\mathbf{x}) + L(L_y)_y(\mathbf{x})) \tag{3.10}$$

The trace and determinant include a normalization component σ_I, the scale of the current pyramidal level as suggested [Lin98]. One particular strength of Hessian-

Laplace is the non-requirement of any thresholding. The algorithm is very similar to Lowe's DOG: The trace approximates DOG and the determinant penalizes edges similarly to thresholding the ratio of Hessian eigenvalues [Low03]. Hessian-Affine, the affine invariant version of Hessian-Laplace, has been found to have the highest interest point accuracy other than MSER [MTS$^+$05].

3.6 Interest point comparison metrics

Two metrics are to evaluate interest points. Repeatability is a binary valued property of an interest point that specifies if it was found invariant to similarity transform. Accuracy measures the degree of invariance of an interest point, and is highly related to repeatability. Algorithms are generally measured in terms of overall repeatability. Accuracy provides a more detailed measure of the quality of an interest point but cannot be computed independently of repeatability.

3.6.1 Repeatability

The optimal selective attention algorithm is invariant to similarity transforms:

$$T(K(I)) = K(T(I)) \tag{3.11}$$

or, equivalently

$$K(I) = T^{-1}(K(T(I))) \tag{3.12}$$

where I is an image, $K()$ is an interest point detector (such as DOG, Harris-Laplace, or Hessian-Laplace), and $T()$ is a similarity transform.

Our algorithm for computing repeatability is as follows. Let $t_i \in T^{-1}(K(T(I)))$ be a interest point from the target image image, transformed back into into the source

image coordinates, and let $t_{i\sigma}$ be the scale of t_i. Similarly, let $s_j \in K(I)$ be an interest point from the original, unmodified image. Then t_i and s_j match if:

$$\frac{t_{i\sigma}}{2^{1/3}} \leq s_{j\sigma} \leq 2^{1/3} t_{i\sigma}$$
$$|t_i - s_j| \leq max(t_{i\sigma}, s_{j\sigma}) \tag{3.13}$$

This metric determines when a target interest point t_i is considered a repeat by being equivalent to s_j. If the scale difference between t_i and s_j is within one-third of an octave and the distance is less than the larger of the two radii, then the interest points match. If a single target interest point t_i matches two target interest points s_j and s_k, only the match with the smaller spatial (as opposed to scale) distance is used.

3.6.2 Accuracy

Accuracy is used for the matching criteria in three recent comparison papers [MS02, MS01, MTS$^+$05] and measures the overlap of an original interest point with its repeat. Accuracy for each interest point is measured as the inverse of error ϵ_S

$$1 - \epsilon_S = \frac{\pi r_s^2 \cap \pi r_t^2}{\pi r_s^2 \cup \pi r_t^2} \tag{3.14}$$

where πr_i^2 is the area of the source or target interest point. Our repeatability metric is equivalent to accuracy thresholding at ≥ 0.227.

3.7 Implementation differences and discussion

Side-by-side implementation of these three algorithms required some design compromises. In addition to numerous threshold decisions that are avoided entirely in our implementation, structural decisions such as interest point σ, number of levels per octave,

and the criteria for detecting a match vary slightly from the original texts. This section explains a few of those problems and our solutions to them.

Our scale space uses three levels per octave because of Lowe, who found improved repeatability up to $s = 3$ and diminishing returns thereafter. For three levels per octave, $\sigma = 1.26$, which is close to Mikolajczyk and Schmid's recommendation of $\sigma = 1.2$ for their H-L algorithms.

We use $\sqrt[3]{2}$ in our repeatability measure to produce keypoints with stronger matching characteristics. This differs from Draper and Lionelle [DL04] who use a set radius of 17 pixels and Mikolajczyk et al. [MTS$^+$05, MLS05] who require $|t_i - s_j| \leq 1.5$. Our repeatability measure is similar to Lowe [Low03] who allows matches within one-half an octave. Matching interest points up to one-third of an octave follows intuitively from using three levels per octave, allowing matches between neighboring levels only.

Harris and Hessian-Laplace interest points are formed using extrema in the DOG signal, rather than the canonical LOG. Crowley et al. confirm that DOG is a good approximation for LOG, showing that the difference in σ between the two functions is a constant, and that the error between the LOG and DOG methods is minimized at $\sigma_{log} = 1.18\sigma_{dog}$ at 3.6% [CRP02].

Extrema are not thresholded following the systems implementations of the previous authors. In SIFT [Low03], DOG extrema are culled based on the ratio of the eigenvalues of the Hessian $(r + 1)/r$ when $r < 10$. Interest points that have a DOG value below a threshold are also thrown away. In the Harris-Laplace method, interest points are culled when their Harris score R is below a threshold. This threshold from Mikolajczyk's PhD thesis is set to 1000. We assume his use of the same threshold in later work.

Chapter 4

Experiments

We measure the repeatability of 10^6 interest points detected from randomly selected images in the CalTech-101 database [FFFP07]. Three algorithms are used for interest point detection: the DOG, Harris-Laplace, and Hessian-Laplace algorithms. Interest points for each algorithm are generated on the exact same set of images and transformations, showing the relative density, repeatability, and accuracy of each interest point detector in Table 4.1.

The transformation $T()$ used to produce interest points from target images t_i is also randomly selected. One quarter of the transformations are a rotation up to 90 degrees, one quarter undergo uniform scaling from 0.9 to 1.2, one quarter apply a -10% to 10% affine transformation, and one quarter randomly combine all three.

Table 4.1: Initial results verifying expected repeatability rates and interest point density of each algorithm.

type	number	repeatability	accuracy
DOG/LOG	311,149	88 %	0.69
Harris-Laplace	1,112,983	85 %	0.68
Hessian-Laplace	381,910	85 %	0.72

Based on extrema detection in the absence of any thresholding or local attribute evaluation, Harris-Laplace produces three times as many interest points as the other algorithms. We therefore randomly select a subset of Harris-Laplace interest points, limiting

the total number of interest points to 10^6 in the following experiments. Repeatability is almost equal across all three algorithms, and the mean accuracy of Hessian-Laplace is slightly higher than either DOG or Harris-Laplace, which confirms previous affine comparison results [MTS$^+$05].

We seek to model the repeatability of individual interest points in the following sections. Section 4.1 describes the attributes we extract at each interest point location. The thresholding decisions of prior authors are confirmed and verified in Section 4.2. We apply the predictions of a generalized linear model (GLM) in Section 4.3 to attribute normalization in Section 4.3.1, the predictability of each algorithm separately in 4.3.2, and to the contribution of each attribute to interest point predictability in general (Section 4.4). Section 4.5 examines the difference between interest points located at minima and those located at maxima. In Section 4.6, we discuss an unexpected and interesting effect of the method used to select neighborhood extrema.

4.1 Attributes of interest points

Seventeen attributes are recorded from each interest point. Each attribute comes from one of five feature "families" that are based on interest point detection algorithms. Regardless of which algorithm detected a specific interest point, attributes are recorded from every feature family. The five families of attributes are position, Harris, Hessian, value, and entropy.

- Position attributes include $xpos$, $ypos$, and $zpos$. $xpos$ and $ypos$ are rescaled to be in a range from 0 to 1.0 relative to their source image dimensions. We do not expect $xpos$ and $ypos$, the x and y coordinates of an interest point in the original image, to have a great effect on repeatability. The scale attribute of an interest point is recorded by $zpos$ and will be informative.

- Harris attributes include $harlambda1$, the first eigenvalue of the second moment

matrix, $harlambda2$, the second eigenvalue of the second moment matrix, and $hardeterminant$, their product. Harris interest points are maxima of R, used in Harris-Laplace.

- Hessian attributes include $heslambda1$, $heslambda2$, and $hesdeterminant$, where $heslambda1$ is the first eigenvalue of the second derivative matrix, $heslambda2$ is the second eigenvalue, and $hesdeterminant$ is their product. Hessian interest points are maxima simultaneously of $hesdeterminant$ and $heslambda1 + heslambda2$.

- A number of successful interest point detectors use entropy [Gil98, KB01, KZB04]. The interest point detectors used in this study are derivative based, rather than entropy, but we include an entropy measure because of its relevance to interest point research. In this family we include $entropy$, the entropy $\mathcal{H} = -\sum p(x) log_D p(x)$ of the region defined by each interest point. Also included are $dentropy$ and $ddentropy$, the first and second derivatives of the local entropy.

- Value attributes include $value$, $truevalue$, $dx2$, $dy2$, and $dz2$. The $value$ attribute changes depending on which algorithm produced an interest point. For DOG interest points, $value = D(x, y, \sigma)$. For Harris-Laplace interest points, $value = R$, and for Hessian-Laplace interest points $value = \mathrm{DET}(H)$. Our choice of $\mathrm{DET}(H)$ follows from the linear modeling of a GLM. Each interest point receives sub-pixel optimization according to Lowe such that $truevalue = D(x)$. Computing $D(\mathbf{x})$ provides us with a 3D quadratic, from which we compute $dx2$, $dy2$, and $dz2$. These are the second derivatives of $D(\mathbf{x})$ in the x,y, and z direction. These features describe local curvature around each interest point, regardless of algorithm.

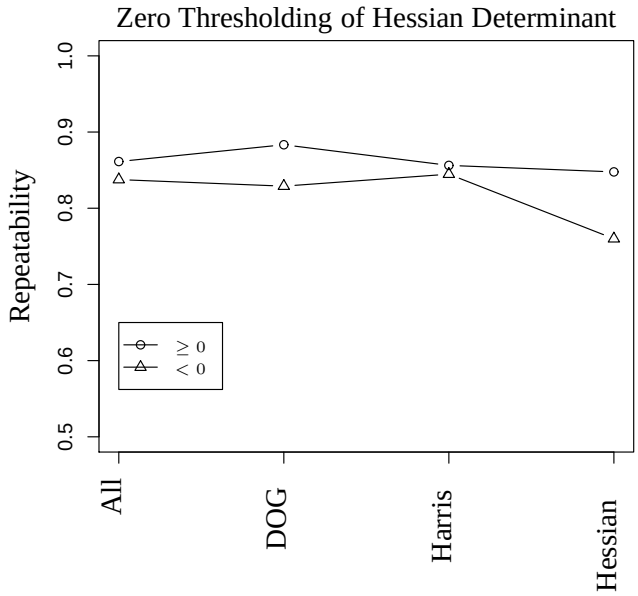

Figure 4.1: Repeatability of interest points thresholded by the Hessian determinant as suggested by Lowe [Low03]. Repeatability is maximized by discarding interest points with a negative Hessian determinant.

We begin our attempt at predicting interest point repeatability by reproducing the thresholding tests of previous authors in Section 4.2. This focuses on a small subset of the available attributes. In Sections 4.3 and 4.4 we attempt to predict interest point repeatability using logistic regression on each attribute.

4.2 Attribute thresholding

The only technique used for improving repeatability in the original works depends on discarding interest points for which a certain attribute falls outside of a threshold. DOG attributes are discarded if the absolute value of the 3D quadratic equation $D(\mathbf{x})$ is below a threshold and if the ratio r of the first and second eigenvalues of the Hessian is greater than ten. Harris-Laplace interest points are discarded if the absolute value of the Harris measure $R < 1000$.

Lowe [Low03] suggests that when the determinant of the Hessian H is negative

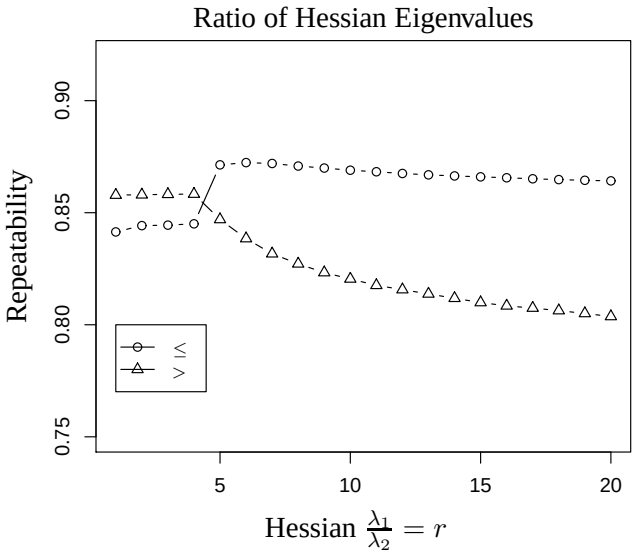

Figure 4.2: Relationship of repeatability and ratio of Hessian eigenvalues. Repeatability is maximized for interest points where $r \leq 5$ regardless of the algorithm used for detection.

the interest point should be discarded. These interest points where the first and second derivatives have opposite signs are edge-like troughs or ridges, instead of peaks. This implies that such interest points will be less repeatable and is well supported by our results. Figure 4.1 shows the repeatability of interest points from each algorithm when thresholded by the sign of the Hessian determinant. 14% of DOG interest points have negative Hessian determinant. These interest points have 83% repeatability and the points with positive Hessian determinants are 88.3% repeatable. Only two percent of the determinant of Hessian-Laplace interest points and almost 40% of Harris-Laplace points are below 0. If an application depends on a small number of highly repeatable interest points, discarding Harris-Laplace points according to this threshold is recommended.

We also test the repeatability of interest points for a range of $r = \frac{heslambda1}{heslambda2}$ to verify Lowe's use of $r < 10$. We find that $r < 5$ is most repeatable for DOG, Harris-Laplace, and Hessian-Laplace interest points. Figure 4.2 show the results of validating Lowe's $r < 10$ threshold [Low03].

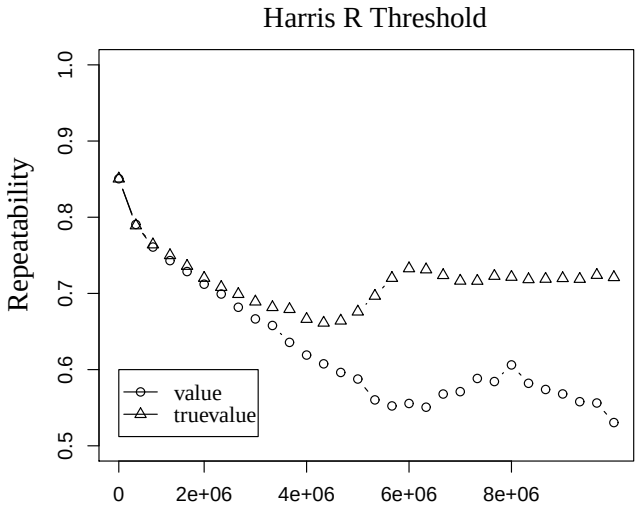

Figure 4.3: Repeatability of interest points with Harris R values above a threshold. We did not find a Harris threshold that improves repeatability. These results show that performance decreases as R increases. We also examined the accuracy of Harris interest points with a similar result.

Harris interest points in prior work are those that are above $R = 1000$ [Mik02]. Figure 4.3 shows the results of tests on the minimum acceptable Harris-Laplace $value$ scores in order to determine that threshold. Included are thresholds against the original $value$ produced from Harris-Laplace, and the sub-pixel optimized $truevalue$ produced from fitting the local region of the interest point to a 3D quadratic equation. We find that no such threshold exists, and that interest point repeatability decreases as R increases. This experiment was also performed using the accuracy metric with identical results.

4.3 Logistic regression

The unique effort of this research is to contribute to runtime prediction of keypoint stability. We investigate this objective through logistic regression on the attributes in Section 4.1 at each interest point. The interest point algorithms being compared each select repeatable locations based on two measures: selecting the maxima of their cor-

responding function (DOG, R, or H) and by thresholding those extrema according to some minimum. This suggests that larger values of these functions are better. We test this hypothesis with the use of a generalized linear model trained on individual interest point attributes.

A GLM iteratively computes the expectation or log odds ratio $E(Y)$ of each dependent variable using maximum likelihood such that

$$E(Y) = g(\beta_0 + \sum \beta_j X_j) = \frac{e^{\beta_0 + \sum \beta_j X_j}}{1 + e^{\beta_0 + \sum \beta_j X_j}} \tag{4.1}$$

using the logit link function $g(p_i) = \log(\frac{p_i}{1-p_i})$ and fitting the prediction variable Y (repeat or non-repeat) to a binomial distribution. X is our dataset of 10^6 interest points with 17 attributes, β_i are the coefficients that model a linear relationship between an attribute and the probability of repeat, and Y is the known repeatability of each interest point as computed with the metric from Section 3.6. Logistic regression allows us to predict the probability that an individual interest point will repeat. We measure the effectiveness of each logistic regression experiment using correlation $r_{E(Y),Y}$ and the area-under-curve (AUC).

The correlation between two vectors $E(Y)$ and Y is defined as

$$r_{E(Y),Y} = \frac{\sum Y_i E(Y_i) - N\bar{Y}\bar{E}(Y_i)}{(N-1)S_Y S_{E(Y)}} \tag{4.2}$$

where S_i is the standard deviation of the set.

$r_{E(Y),Y}$ is maximized when the set of $N = 10^6$ samples and the set of expectations $E(Y)$ vary simultaneously. This metric depends heavily on the dimensionality of the data - one-tailed significance for $p < 0.05$ when $N = 10^6$ requires a correlation score of only 0.001645.

Area under the curve measures the discrimination of each fitting to correctly predict an interest point that either repeats or does not repeat. It is the measured area under a

receiver-operating-characteristic (ROC) curve which plots a fitting's $sensitivity$ against $1 - specificity$.

Because interest points are generated from three algorithms that may produce attributes with different variance, we first examine data normalization techniques in Section 4.3.1. Section 4.3.2 tests if either of the three selective attention algorithms is more or less predictable via GLM. Section 4.4 closely examines the performance of our GLM on each family of interest points. We find, ultimately, that the repeatability of an individual interest point cannot be easily predicted using a linear model. We see that each attribute influences repeatability, but none strongly. This will enable us to construct a GLM that increases repeatability 4% by ranking the interest points from most-to-worst likely to repeat.

4.3.1 Normalization techniques

Attributes of the original data have variances that range from slightly above zero to 10^{11}. This variability suggests that we look initially at normalization techniques. We investigate three normalization techniques on the data including mean centering each sample and giving it unit length

$$unit(X_{ij}) = \frac{X_{ij} - \bar{X}_i}{\sqrt{\sum X_i^2}} \tag{4.3}$$

Mean centering each sample and dividing the attributes X_j by their standard deviation

$$sd(X_{ij}) = \frac{X_{ij} - \bar{X}_j}{\sqrt{\frac{1}{N}\sum (X_j - \bar{X}_j)^2}} \tag{4.4}$$

Log-normalization of the absolute value of the subset of attributes with the largest variance is also examined including the Harris, Hessian, and value attributes.

$$log(X_{ij}) = log_2(|X_{ij}|) \tag{4.5}$$

Normalization has very little effect on either $r_{E(Y),Y}$ (Figure 4.5) or AUC (Figure 4.4) scores. There is one exception: log normalization of the data affects the correlation of Harris and Hessian fittings. Correlation scores for many of these attributes are inverted. The net positive effect after log normalization is that the AUC and correlation scores are highly correlated ($r = 0.97, p < 0.0000001$). Before log normalization correlation between our two metrics is $r = 0.474, p = 0.06$. We discuss the meaning of this behavior in Sections 4.4.2 and 4.4.3.

The only cases where fitting with original attributes underperforms any normalization technique are $hardeterminant$ and $truevalue$. In both cases, the original model is deceived by extremely large outliers which are corrected for by log normalization. We believe that the correlation and AUC metrics are effective because log normalization introduces such a strong correspondence between them. Original attributes and log normalized attributes are superior to unit and sd normalization in every fitting. We proceed into Section 4.3.2, GLM performance by algorithm, and Section 4.4, GLM performance by each specific attribute, with an investigation of the effect of original and log normalized attributes on repeatability.

4.3.2 Regression by interest point detector

Figures 4.6 and 4.7 show the performance of logistic regression on the original data and its log, separated by which algorithm the interest point was detected by. Interest points generated from the three algorithms do not, as we expected, depend primarily on their own attributes to predict repeatability. Nearly every attribute has some predictive power across all three algorithms; however that predictive power is quite weak. A horizontal line is drawn at the random line for AUC $= 0.5$. Our predictions have $999,998$

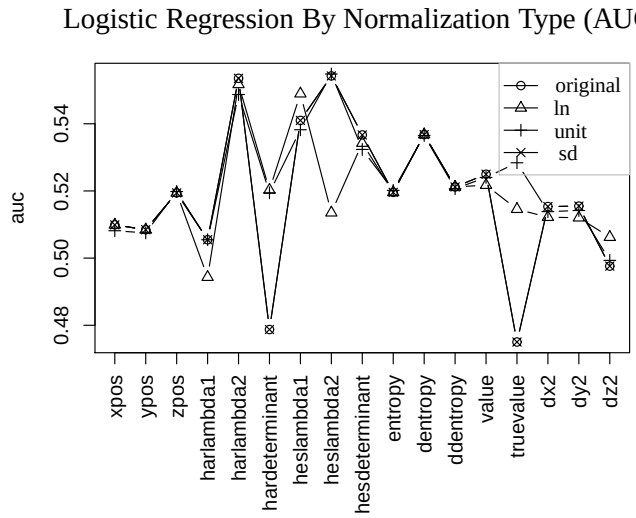

Figure 4.4: Logistic regression by normalization type. The area under the curve (AUC) for each attribute on original data and each of three common normalization techniques. Fitting determinant of Harris $hardeterminant$ and optimized value $truevalue$ attributes fail because of the magnitude of these attributes.

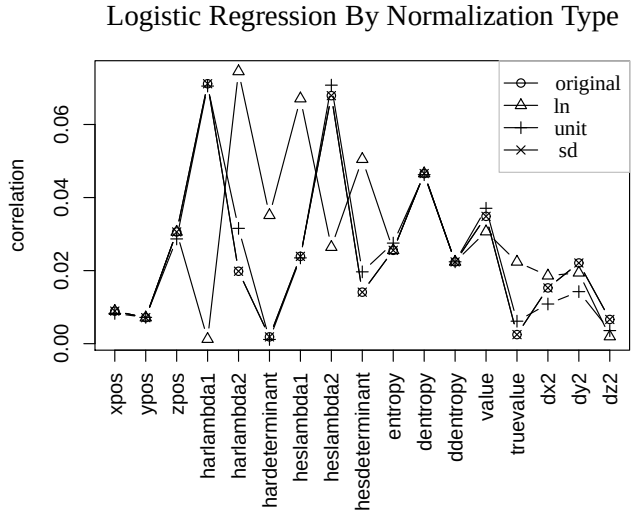

Figure 4.5: Logistic regression by normalization type. The correlation $r_{E(Y),Y}$ for each attribute on original data and each of three common normalization techniques. Log normalization reduces quality of fit for only two attributes and causes $r_{E(Y),Y}$ and AUC to correspond highly ($p < 0.0000001$).

degrees of freedom, so a significance of $p = 0.05$ in a one-tailed t.test is achieved at $r = 0.001645$. A horizontal line is also drawn at that point on the graph, nearly indistinguishable from zero.

The $harlambda1$ attribute produces the largest $r_{E(Y),Y}$ but does not boost AUC because AUC is a rank-based metric. The steep slope of the GLM seen in Figure 4.11 is enabling the prediction to produce a larger set of probable repeats without properly ranking them. A similar effect is causing high correlation with $heslambda2$, visualized in Figure 4.19.

Results for why AUC is maximized among Harris-Laplace interest points using a GLM fit to Hessian family attributes is unclear. Similar correlation results for DOG with log normalized Hessian attributes is less surprising, as both are blob detectors.

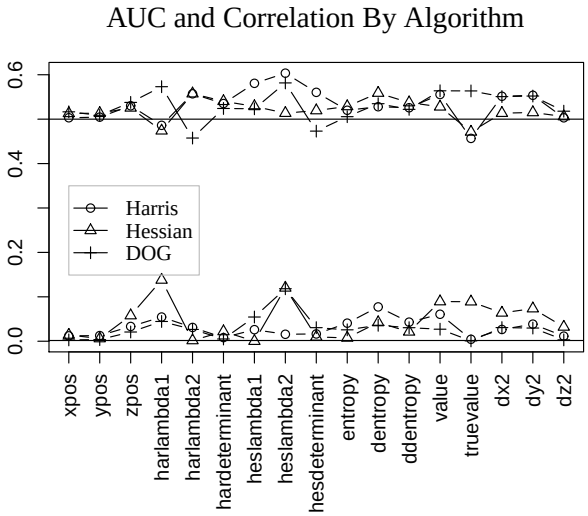

Figure 4.6: Regression by algorithm on original attributes. Performance of the GLM predictions are low but non-random. The fitting of Hessian interest points maximizes $r_{E(Y),Y}$ and the Harris fitting maximizes AUC.

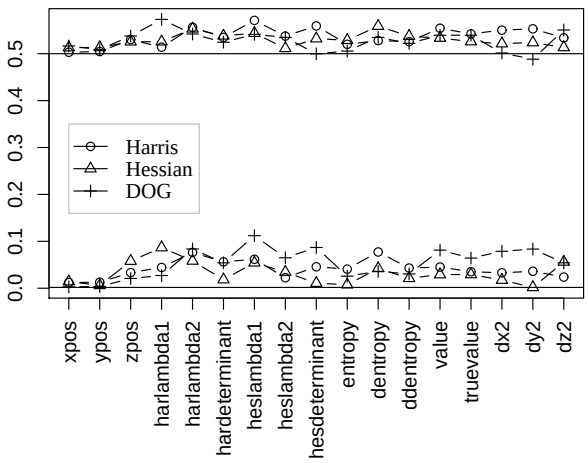

Figure 4.7: Regression by algorithm on log attributes. Log normalization introduces uniformity and indicates DOG is most predictable.

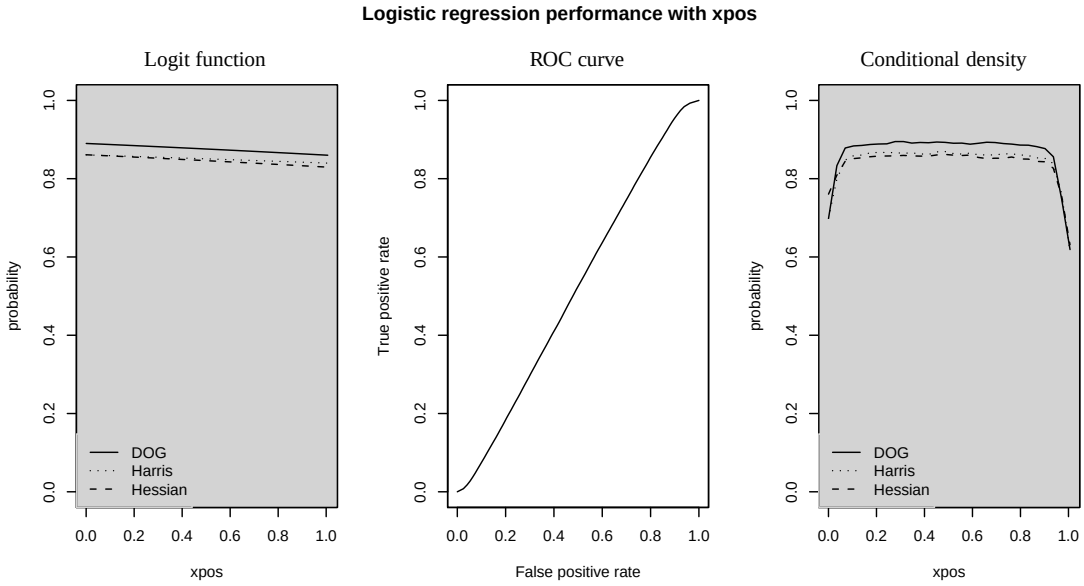

Figure 4.8: Extremum near the borders of images are predictably not as repeatable. $r_{E(Y),Y} = 0.01, AUC = 0.51$

4.4 Individual attribute performance

The next section shows a set of three graphs for each interesting attribute with the goal of finding shared attribute dependencies among all three algorithms, and exploiting them. Each graph includes the logit probability function produced by the GLM from the indicated feature, the ROC curve of the logit prediction, and a conditional density graph showing probability of a repeat by attribute value. A grey box is drawn on the logit function and conditional density plot denoting the boundaries of two standard deviations above and below the mean of samples for each attribute.

The logit function shows the potential strength of the prediction. Uninformative attributes like the image coordinates of an interest point appear flat. More informative attributes will actually appear as a logit function with a distinct boundary between the probability of a repeat and a non-repeat. However, none of our attributes are strongly predictive of a repeat or non-repeat. Most of the logit functions are approximately linear,

particularly within two standard deviations of the mean. The logit functions with a high positive or negative slope have some predictive effect on repeatability and suggest thresholding or weighting interest points by these features.

The ROC curve provides a visualization of the AUC score from sections 4.3.1 and 4.3.2. It graphs the $sensitivity$ of a classifier against $1 - specificity$. Sensitivity is the number of true positives T_p over the sum of true positives T_p and false negatives F_n. Specificity is the number of misclassified negatives: $F_p/(F_p + T_n)$. A ROC curve for a classifier with random performance is a line with slope $= 1$ and AUC$= 0.5$. None of the AUC scores are above 0.6 and none of the ROC curves appear to be very strong classifiers, but they are a visual aid to the performance of each logistic regression. As the curve stretches toward the top left corner (perfect sensitivity and specificity) the fitting is more predictive.

The conditional density of an attribute is given by $p(y|x_i)$ where x_i is a particular range of values of the attribute x. It is computed from Bayes rule as

$$p(y|x) = \frac{p(y)p(x|y)}{p(x)} \tag{4.6}$$

and in a discrete sample is simply

$$\frac{\sum_i p(y|x_i)}{p(y)} \tag{4.7}$$

over each attribute. This graph shows the attribute range where repeatability is maximized. It is particularly informative with log normalized attributes and suggests a number of thresholding decisions.

Figure 4.8 shows these three graphs based on the interest point x, y coordinates. As expected, interest points near the border of an image demonstrate reduced repeatability caused by border effects. Otherwise, the coordinates of an interest point have no effect on its repeatability. We suggest discarding interest points within one-twentieth of the

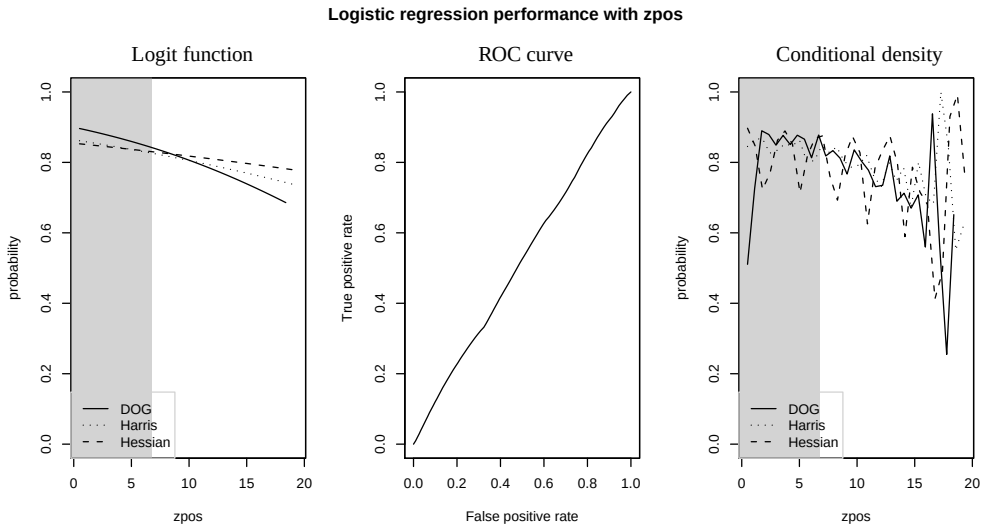

Figure 4.9: Logit function predicted by the GLM, ROC curve, and conditional density estimation of scale. Hessian points are the most stable to scale increases, with the most stable points at the bottom of an octave and the least stable at the top. $r_{E(Y),Y} = 0.03, AUC = 0.52$

image border.

4.4.1 Scale

Interested in the effect of σ on repeatability, we show the repeatability of each algorithm as a function of scale in Figure 4.9. Hessian-Laplace produces highly repeatable interest points on every octave, though interest points on the highest level of each octave have consistently low repeatability. We believe this is because of the σ-normalization of derivatives, which is too large for interest points normalized by σ_I^2. Interest points on the bottom level of each octave are normalized by $1/2^{1/3} < 1$, giving them a lower likelihood of being maximal. Therefore only the most stable interest points remain extremal after normalization, increasing the stability of the bottom octave. DOG fitting produces similar though less pronounced results.

The repeatability of every algorithm is sinusoidal and is dependent on the production

of the scale space. DOG interest points seem to repeat most often in the center level of the octave where there are a set of 26 neighbors in a cube around them. Harris interest points behavior in a similar fashion, preferring octaves with a full neighborhood. Hessian interest points repeat the least often in the middle of an octave, however. We suspect that this is a function of the neighborhood operator discussed in Section 4.6.

Repeatability of interest points at the highest levels is unpredictable because of insufficient samples. The number of interest points decrease logarithmically with scale because image size is quartered at each octave. We suggest that researchers utilizing interest point detection who desire a small number of repeatable interest points select only those interest points with a moderate level of scale. Repeatability decreases only slightly (particularly with Hessian-Laplace) and density decreases significantly.

4.4.2 Harris eigenvalues

Harris eigenvalues are seen in six figures: 4.11, 4.12, 4.13, 4.14. 4.15, 4.16. Correlation scores using these features are inverted with log normalization in Figure 4.5 because of variance reduction. The first eigenvalue of the Harris and its determinant have the largest variance of any feature: $\sigma^2 = 10^6$ and 10^{11}, respectively. Log normalization reduces this and with it the GLM's tendency to overfit.

The first eigenvalue of the Harris matrix $harlambda1$ shows the highest correlation of any feature (see Figure 4.6), suggesting the relationship between $harlambda1$ and repeatability is linear. This is true on the original features where the fit has a large negative slope. Logging introduces nonlinearity to the feature, eliminating the strong negative slope and reducing fit performance. Interest points with a large $harlambda1$ are edges and should always be discarded.

It is easy to understand why the original GLM prediction of $harlambda2$ is the inverse of $harlambda1$. Repeatability is maximized for interest points with small $harlambda1$ and large $harlambda2$ implies that repeatability is maximized when their

ratio is minimized. The result of minimizing their ratio is seen in Figure 4.10.

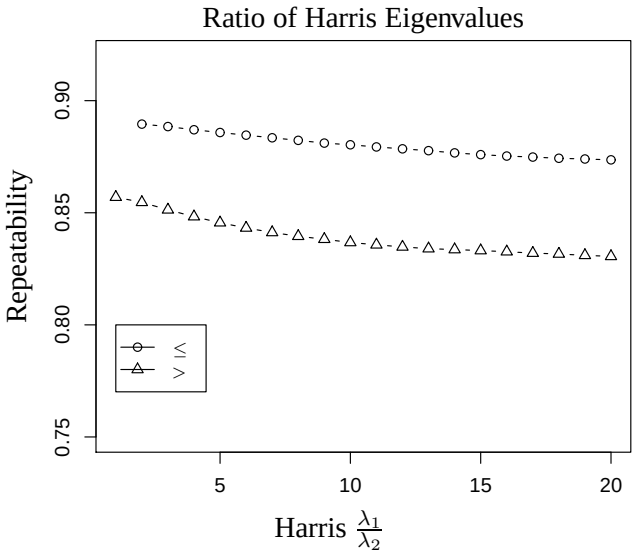

Figure 4.10: Investigation of repeatability of interest points against the ratio of Harris eigenvalues. Our results show repeatability of almost 90% for interest points with eigenvalue ratio below 5.

The determinant of the second moment matrix, $hardeterminant$ does not predict repeatability well. This may seem unintuitive since Harris-Laplace uses this value to select interest points. The large variance of $harlambda1$ reduces the informativeness of $hardeterminant$, which is the product of the two eigenvalues. The determinant can, it seems, be as easily maximized on an edge as on a corner. As Figure 4.10 demonstrates, the ratio of Harris eigenvalues contributes most to repeatability.

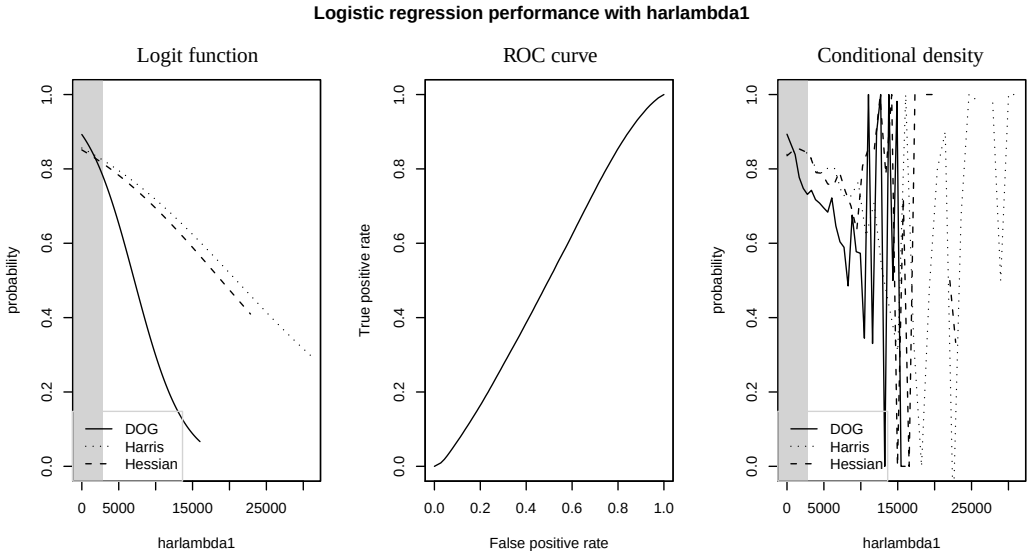

Figure 4.11: Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $harlambda1$. Repeatability decreases as the first eigenvalue increases as interest points become more like edges and less like corners. High correlation and low AUC suggest a bad fit: $r_{E(Y),Y} = 0.07, AUC = 0.51$

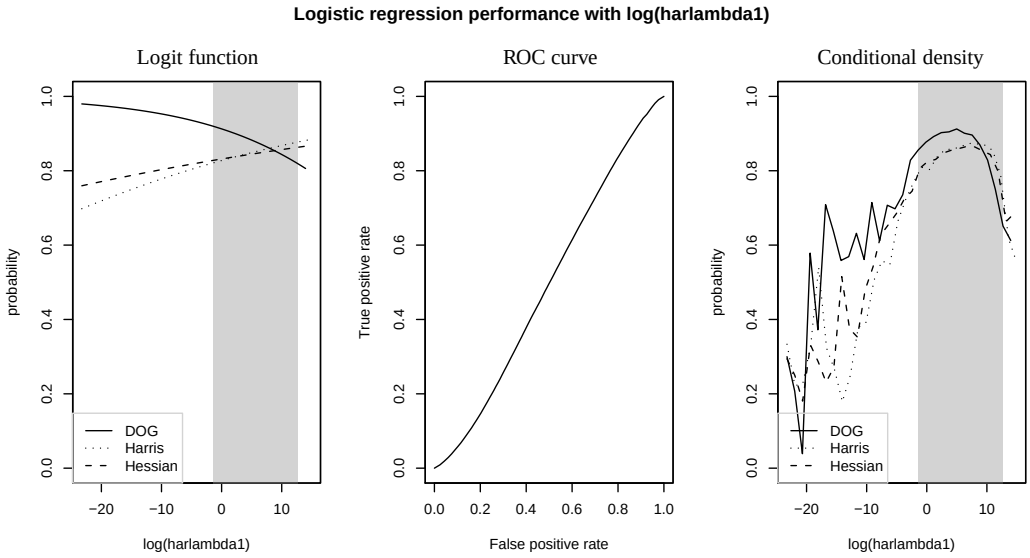

Figure 4.12: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $harlambda1$. $r_{E(Y),Y} = 0.00, AUC = 0.49$

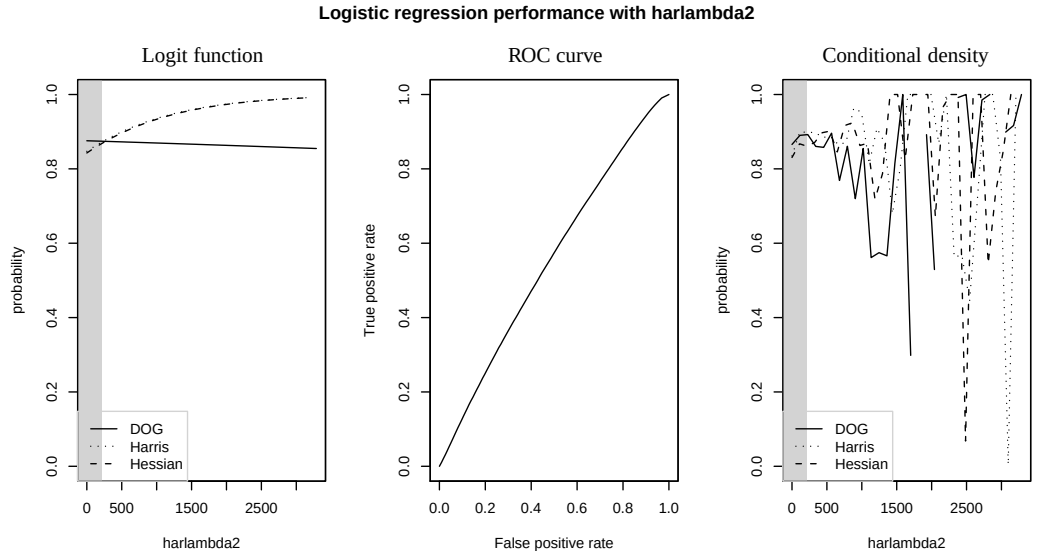

Figure 4.13: Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $harlambda2$. High AUC and low correlation suggest overfitting of the model: $r_{E(Y),Y} = 0.02, AUC = 0.55$

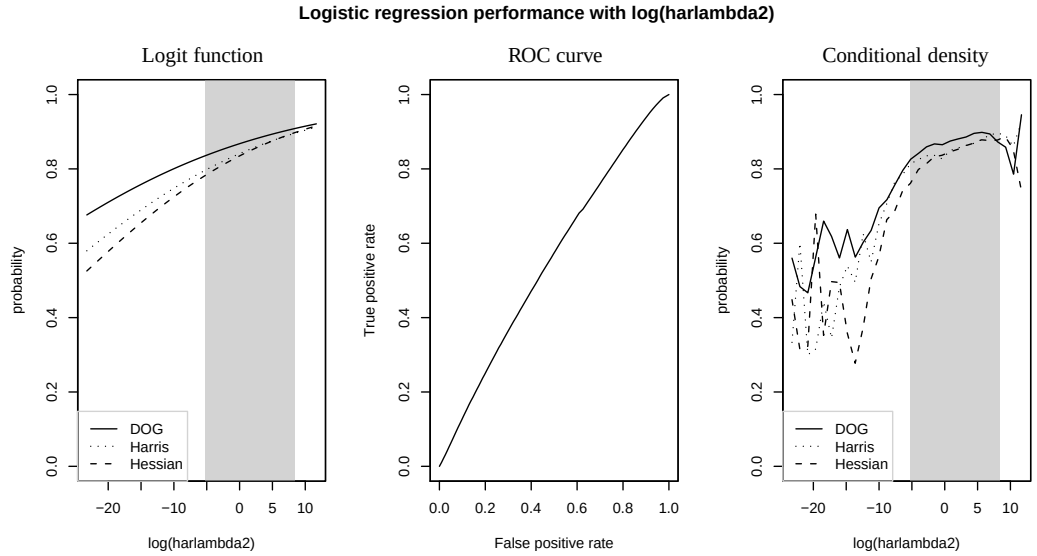

Figure 4.14: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $harlambda2$. Collapsing the variance reveals a clear linear relationship for all three algorithms. The most predictive attribute: $r_{E(Y),Y} = 0.08, AUC = 0.55$

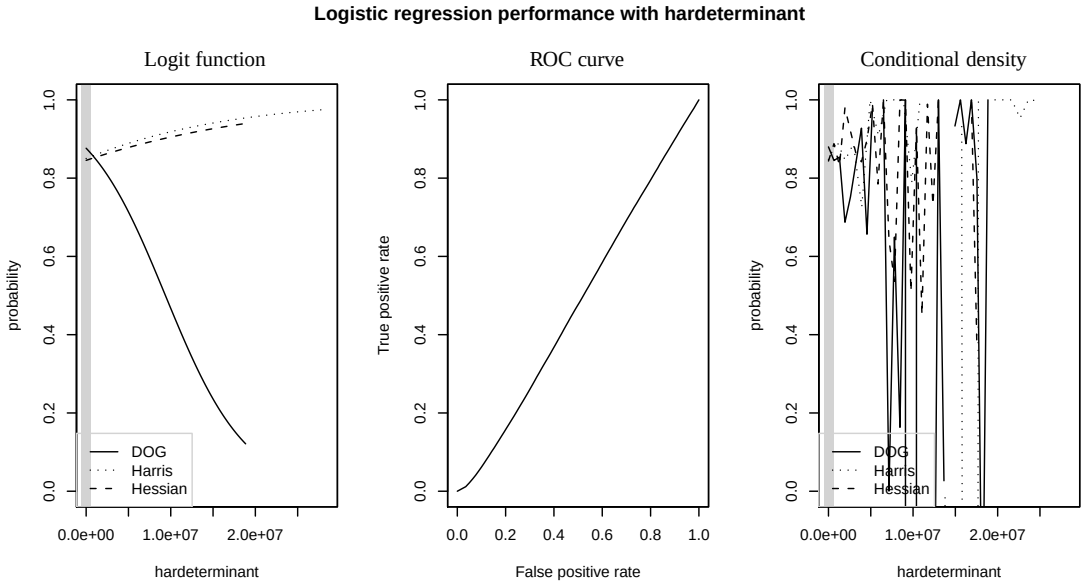

Figure 4.15: Logit function predicted by the GLM, ROC curve, and conditional density estimation of $hardeterminant$. The large difference in slope for DOG is caused by variance, seen in the next figure. $r_{E(Y),Y} = 0.0018, AUC = 0.48$

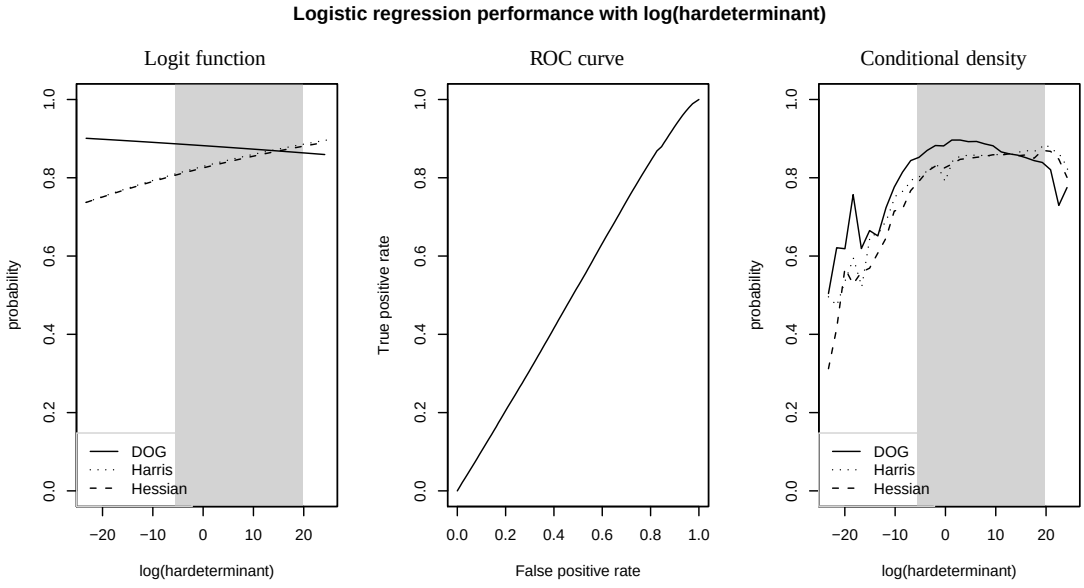

Figure 4.16: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $hardeterminant$. $r_{E(Y),Y} = 0.04, AUC = 0.52$

4.4.3 Hessian eigenvalues

The Hessian matrix H of second derivatives is a blob detector. Prediction performance of this attribute family, along with the Harris attribute family, inverts when log normalization is used. The next six figures, Figures 4.17, 4.18, 4.19 4.20, 4.21, and 4.22 show the results of our experiments on the original and log-normalized values of these attributes.

Regression on the first and second eigenvalues of the Hessian behaves similarly to those of the Harris. The slope of $heslambda1$ is flat or slightly positive while $heslambda2$ is sharply positive. This reflects the dependence, suggested by Lowe [Low03] that repeatable interest points have a small ratio between first and second eigenvalues. The result of our thresholding experiment supporting Lowe's results are seen in Section 4.2. We have also found in the previous section that this applies to Harris eigenvalues.

Each attribute in the Hessian family is nonlinearly related to repeatability except for $hesdeterminant$. A positive slope for the H-L algorithms on this attribute suggests that the Hessian does not respond as strongly to edges as the Harris. The determinant is maximized when both eigenvalues increase simultaneously unlike the Harris when the first eigenvalue overtly weights the determinant.

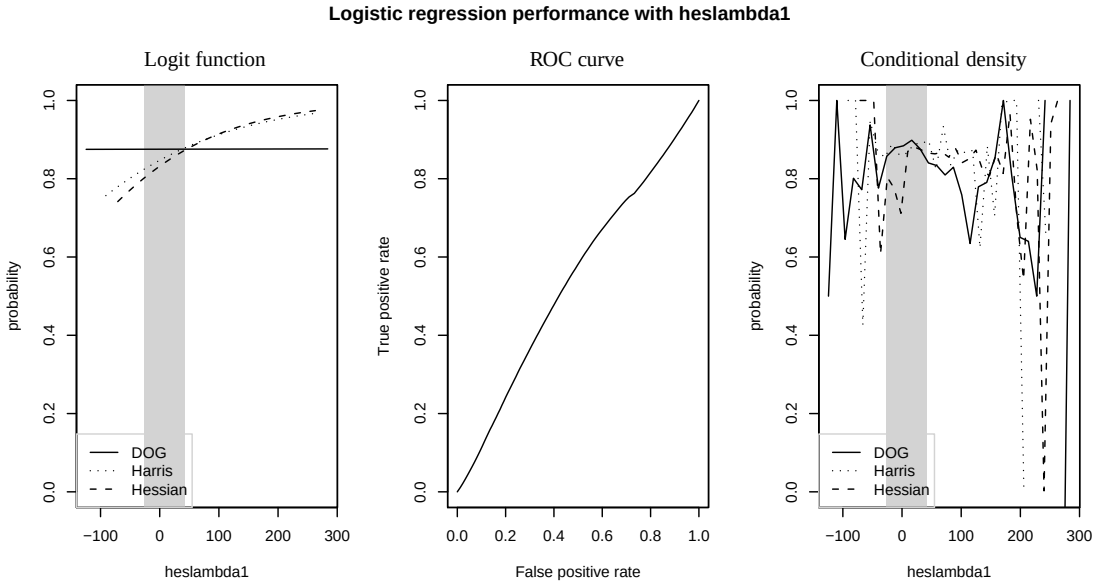

Figure 4.17: Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $heslambda1$. $r_{E(Y),Y} = 0.02, AUC = 0.54$

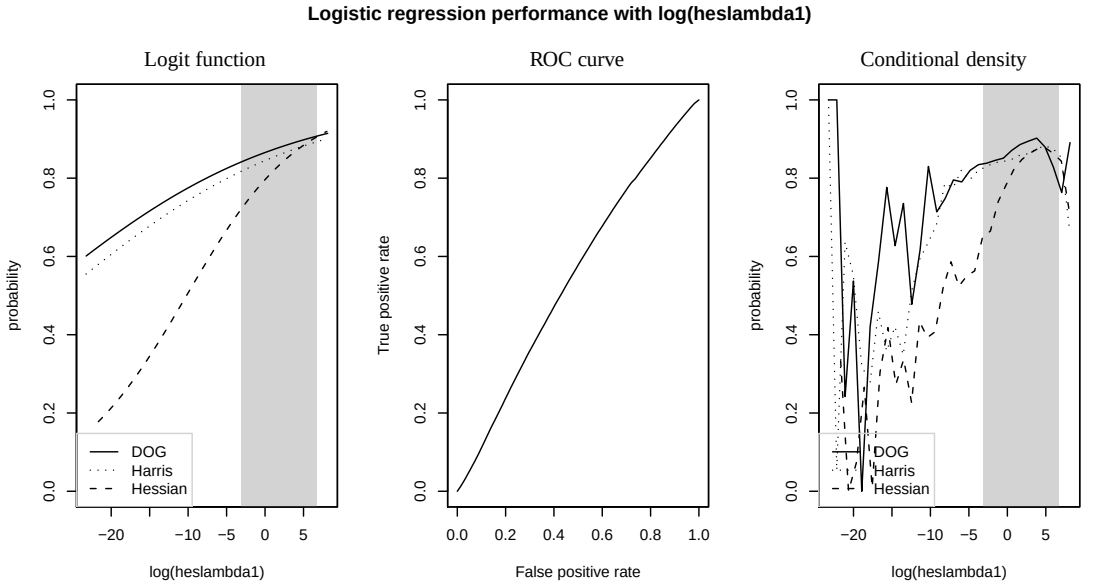

Figure 4.18: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $heslambda1$. $r_{E(Y),Y} = 0.07, AUC = 0.55$

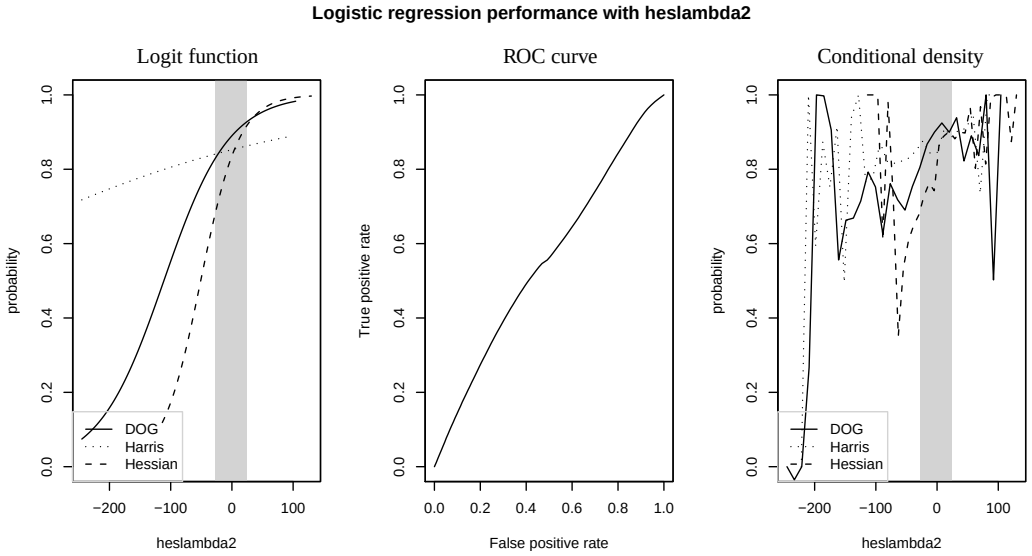

Figure 4.19: Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $heslambda2$. This attribute increases linearly with repeatability and suggests discarding when < 0. $r_{E(Y),Y} = 0.07, AUC = 0.55$

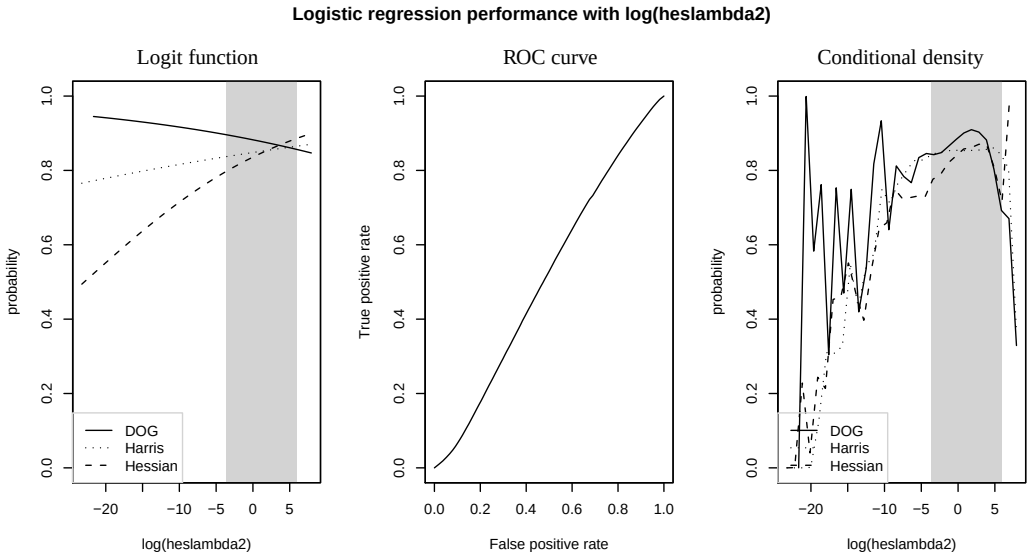

Figure 4.20: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of Hessian $heslambda2$. The strong relationship from the original feature disappears after the absolute value is taken in log normalization. $r_{E(Y),Y} = 0.03, AUC = 0.51$

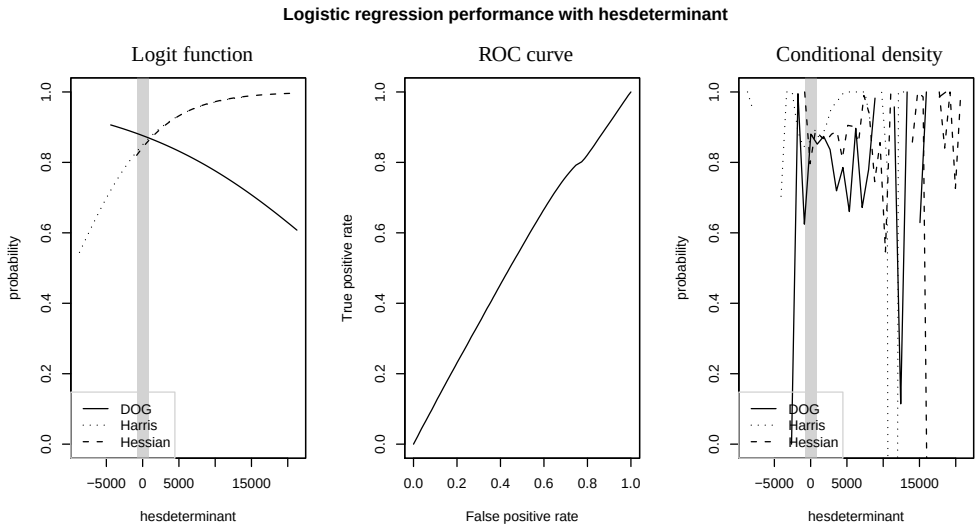

Figure 4.21: Logit function predicted by the GLM, ROC curve, and conditional density estimation of original $hesdeterminant$. The slopes are exaggerated because of high variance and are reduced in the nexture figure. $r_{E(Y),Y} = 0.01, AUC = 0.54$

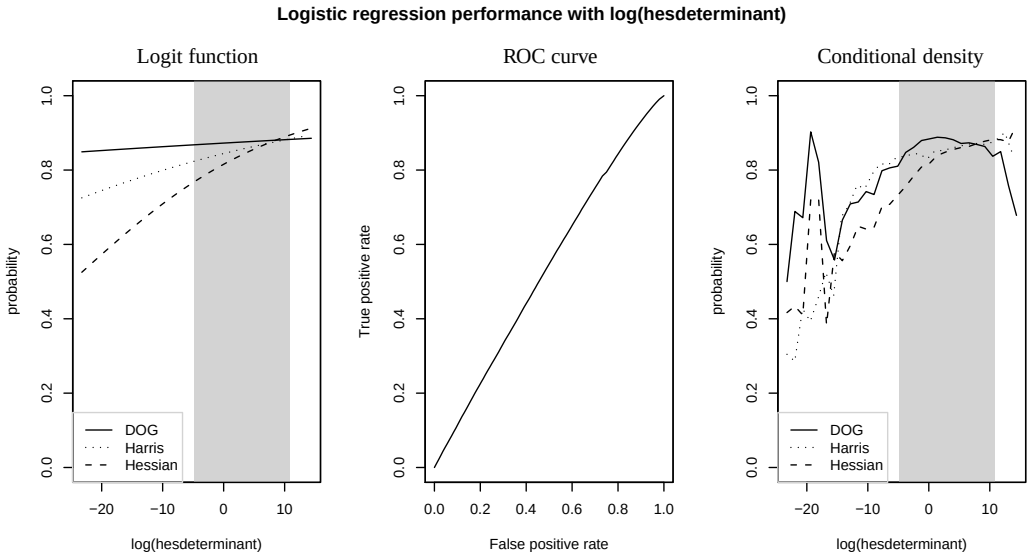

Figure 4.22: Logit function predicted by the GLM, ROC curve, and conditional density estimation of log of $hesdeterminant$. $r_{E(Y),Y} = 0.05, AUC = 0.53$

4.4.4 Entropy scores

Selective attention algorithms have historically appealed to the idea of entropy in detecting stable interest points. We examine the measure of entropy $\mathcal{H} = -\sum p(x) log_D p(x)$ and its first and second derivatives for each interest point. Figures 4.23, 4.24, and 4.25 show the results. Figure 4.4 and Figure 4.5 show that entropy scores are most effective at predicting Harris-Laplace repeatability. Interest points with entropy below 1 in our dataset are 6% less repeatable than others. We see that H-L algorithms' repeatability decreases slightly as the derivative of entropy increases. DOG also benefits from this interpretation of the first derivative.

While the relationship is weak, we see that repeatability does to some degree depend on regions where the rate of change of the entropy measure is decreasing. These regions correspond to a local image region rapidly shifting from light to dark intensity values or vice-versa. Blob and corner regions, as produced by the algorithms in this research, also correspond to this local structure.

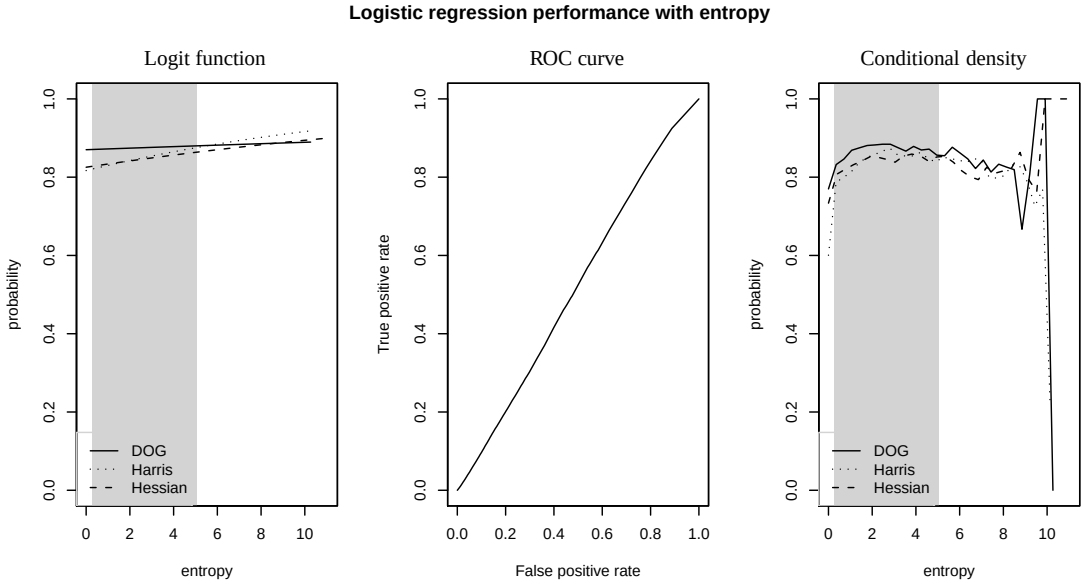

Figure 4.23: Logit function predicted by the GLM, ROC curve, and conditional density estimation of entropy. Interest points with $entropy < 1$ should be discarded. $r_{E(Y),Y} = 0.03, AUC = 0.52$

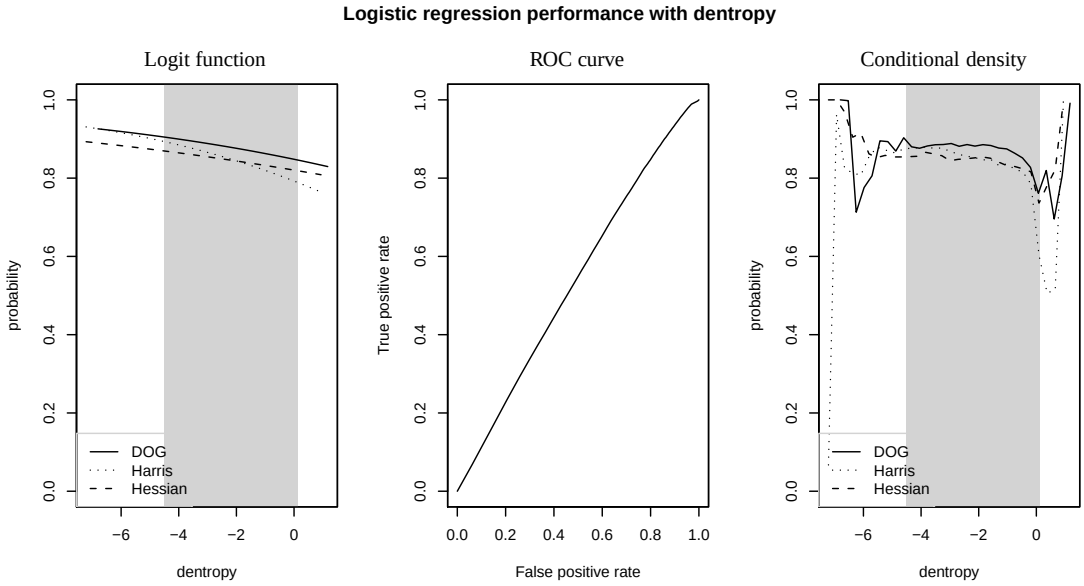

Figure 4.24: Logit function predicted by the GLM, ROC curve, and conditional density estimation of first derivative of entropy. Interest points with $dentropy > -1$ are 4% less repeatable than others. $r_{E(Y),Y} = 0.05, AUC = 0.54$

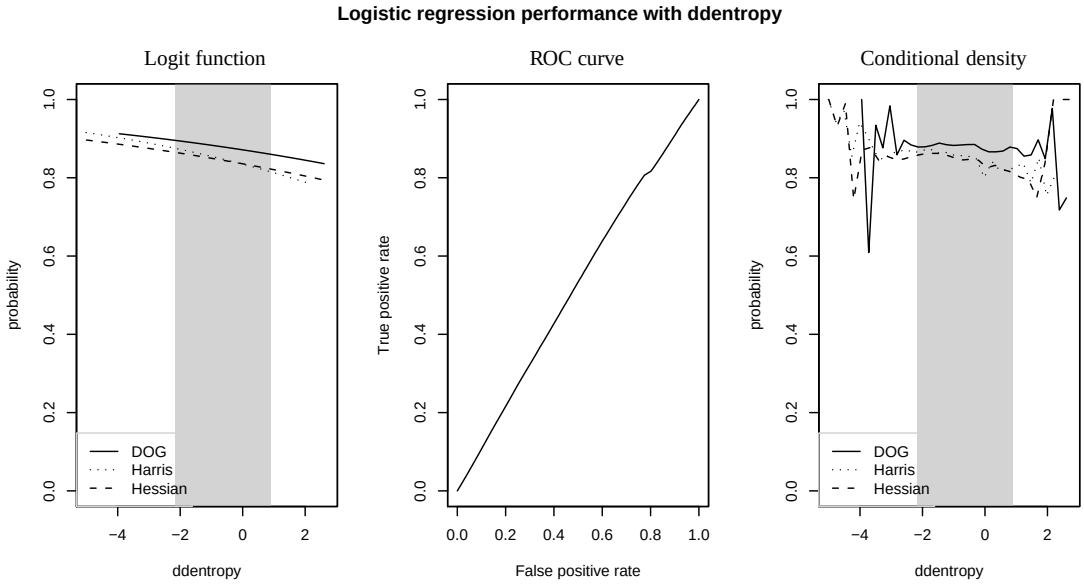

Figure 4.25: Logit function predicted by the GLM, ROC curve, and conditional density estimation of second derivative of entropy. $r_{E(Y),Y} = 0.02, AUC = 0.52$

4.4.5 Values of extrema and their neighborhood

This section examines attributes from the value family. Figures 4.26 and 4.27 plot three separate $value$ scores. For DOG interest points, $value = D(x, \sigma) = L(x, \sigma) - L(x, \sigma - 1)$. For Harris-Laplace $value = R$, and for Hessian-Laplace $value = \mathrm{DET}(H)$. Sub-pixel optimization is performed on $value$ with a 3D quadratic fitting in order to produce $truevalue$. We expect these two measures to be informative in predicting the repeatability of interest points from their source algorithm. Figures 4.28, 4.29, and 4.30 each show the results for the second derivatives of the same 3D quadratic. Log normalized attributes are used in all five figures because of their high variance.

The results are similar for all three algorithms: a large Hessian $value$ is good, DOG $value$ increases repeatability when small, and the Harris $value$ is uninformative. There are two interesting and unexplained results from this experiment. Regression performance in Figure 4.7 shows that DOG is best predicted from these features. The reason is because the variance of DOG values ($\sigma^2 = 92$) in this experiment fall outside of the range of Harris-Laplace ($\sigma^2 = 3.7 \times 10^{11}$) and Hessian-Laplace ($\sigma^2 = 1.3 \times 10^5$) values by a significant margin, biasing the model to take advantage of the slight repeatability advantage of DOG to compute $r_{E(Y),Y}$.

Behavior of the Hessian attributes is unexpected. Each fitting shows a strong positive relationship between each value attribute and repeatability. We were able to increase Hessian-Laplace repeatability to 88% in our experiments by discarding Hessian-Laplace interest points with $value < 1000$.

The derivative attributes $dx2$, $dy2$, and $dz2$ were expected to be most informative because they describe the local neighborhood around an extrema. By definition, an extrema is a local region where the second derivatives in the x and y directions are close to zero. We believe that these measures are ineffective for two reasons. The local region is only descriptive for DOG, which computes the 3D quadratic from a 27-

pixel neighborhood cube of the same function. Harris-Laplace and Hessian-Laplace use an incompatible combination of neighborhood operators, discussed in Section 4.6. The second reason is that while $dx2$ and $dy2$ are computed directly from a quadratic function, that function is oriented along the x and y axes of the original image - not along the principal directions of variance. The second moment matrix from Mikolajczyk and Schmid's affine invariant work [MS02, MS04] suggests another method to compute a 3D quadratic with principal orientation information.

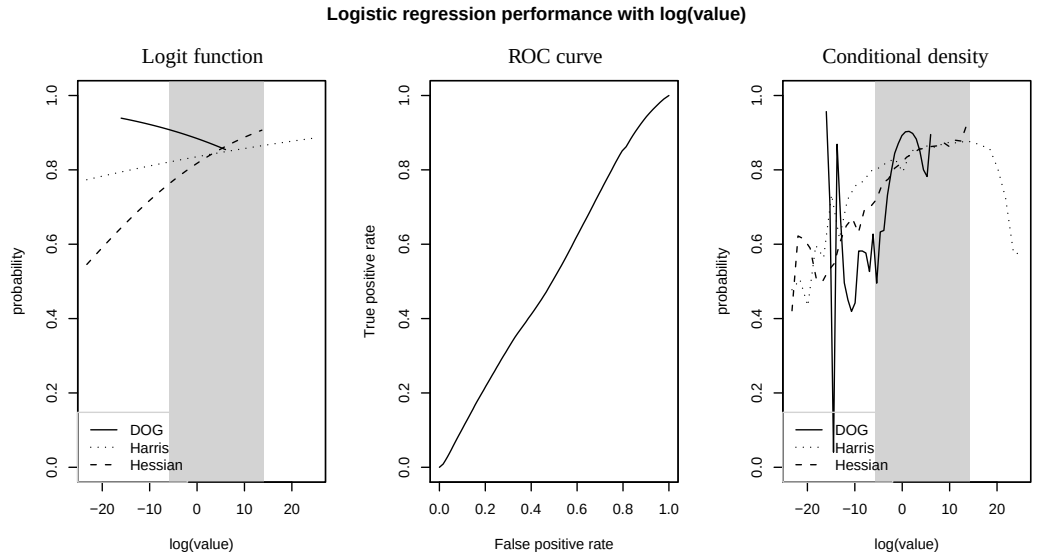

Figure 4.26: Logit function predicted by the GLM, ROC curve, and conditional density estimation of $value$ at each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.03, AUC = 0.52$

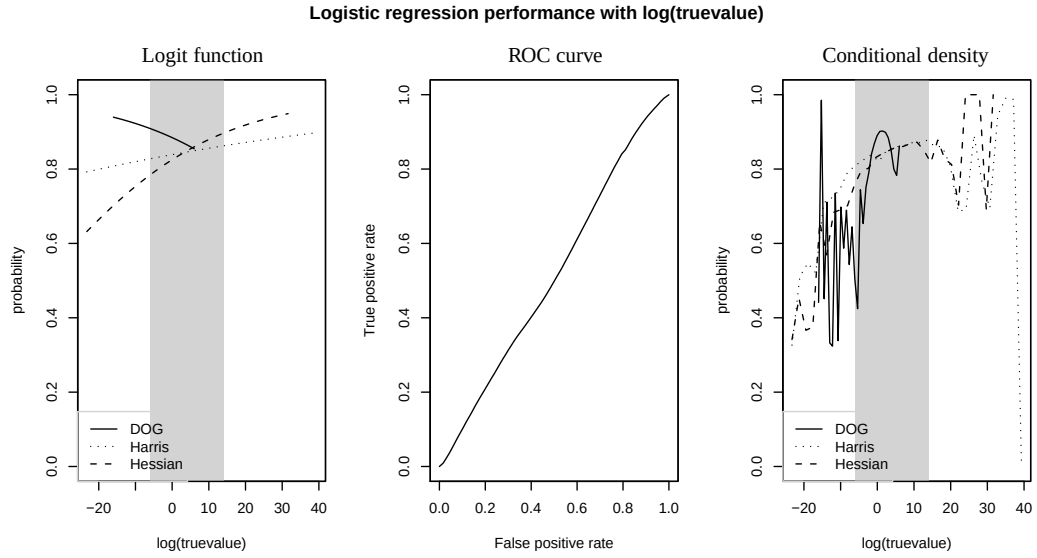

Figure 4.27: Logit function predicted by the GLM, ROC curve, and conditional density estimation of sub-pixel optimized $truevalue$ at each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$

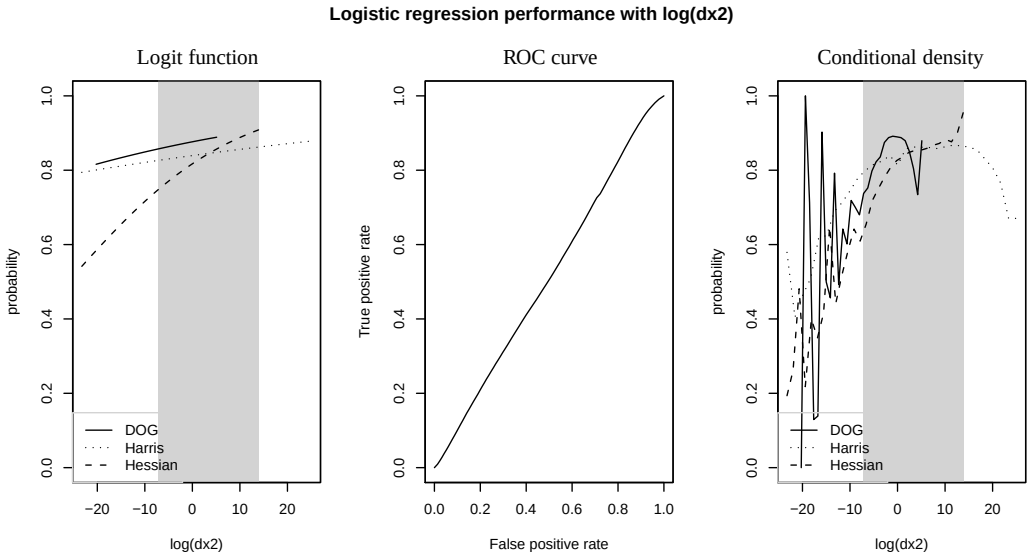

Figure 4.28: Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$

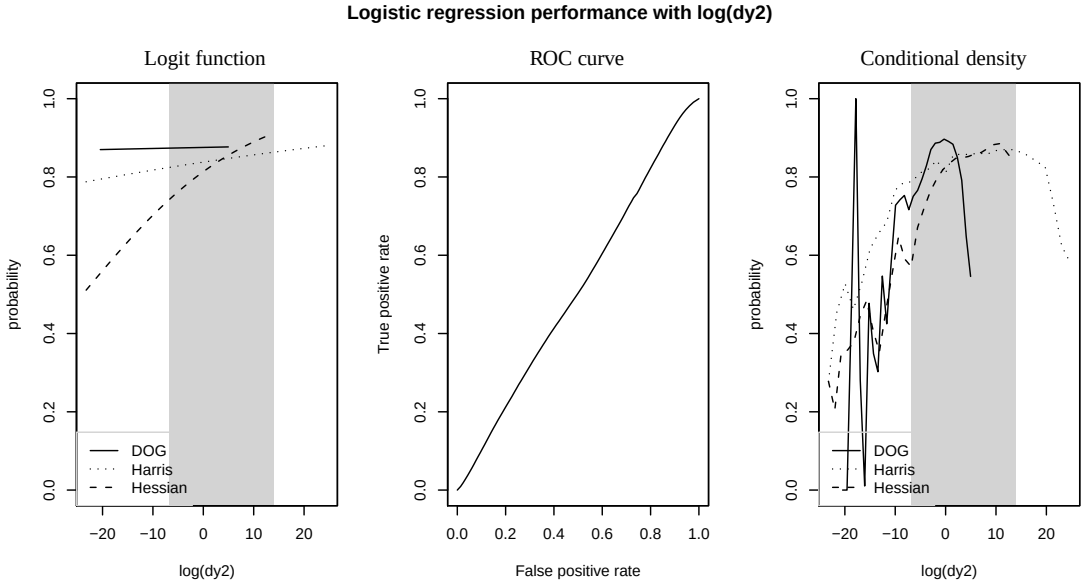

Figure 4.29: Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.02, AUC = 0.51$

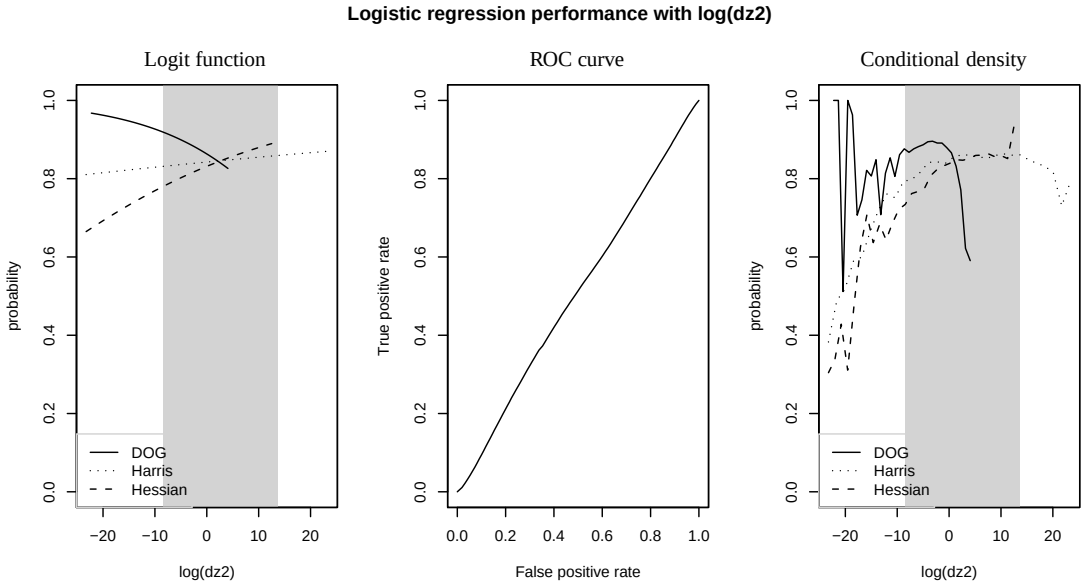

Figure 4.30: Logit function predicted by the GLM, ROC curve, and conditional density estimation of the second derivative with respect to x in the neighborhood of each extremal location ($D(x, y, \sigma)$, R, and $\mathrm{DET}(H)$) $r_{E(Y),Y} = 0.00, AUC = 0.51$

4.5 Extrema inversion

The scale space of an image is a continuous signal containing both maxima and minima. The common technique is to detect only maxima in the absolute value of the scale space signal. This technique converts negative minima to positive maxima but overlooks two additional types of extrema.

Scale space extrema consist of four classes - positive maxima, negative minima, positive minima, and negative maxima. These extrema types are expressed in the four triplets $\{4, 5, 4\}, \{-4, -5, -4\}, \{4, 3, 4\}$,and$\{-4, -3, -4\}$. The second two types of extrema are ignored using the common method of absolute maximum detection. Suspecting this to be an undiscussed repeatability improvement technique, we measured the repeatability of the four classes of extrema in Figure 4.2.

These results verify the implicit suggestion that inverted extrema can be ignored.

Table 4.2: Extrema inversion results. Negatively valued extrema are slightly less repeatable than positive extrema [Low03].

	Overall	+ +	- -	- +	+ -	+	-
repeatability	87 %	88.3 %	85.2 %	61.2 %	58.3 %	88.3 %	85.2 %
number	699,957	343,839	355,349	589	180	344,019	355,938

The repeatability of negative maxima and positive minima is 20% lower and their low frequency suggests they can be ignored completely.

4.6 Method of extrema detection

A scale space presents a discrete representation of the continuous signal of frequences in an image. That representation implies that neighborhoods exist in scale as well as space. Neighborhood extrema can be detected in multiple arrangements. Lowe's DOG approach detects extrema in space and scale simultaneously by only accepting points larger or smaller than their 26 neighbors. If an interest point location is larger than its eight neighbors at the same scale and its nine neighbors above and nine neighbors below it, it is a cube extrema. Harris-Laplace detects Harris extrema in the 8-neighborhood (level neighbors) surrounding the point and ignoring scale neighbors. Characteristic scale localization is performed by simply testing whether the immediate LOG neighbor above and below (tower neighbors) the point are non-maximal. The neighborhood technique for Hessian-Laplace is unspecified. We use a mirror of the Harris-Laplace, detecting level neighbors in the determinant and tower neighbors in the trace. Extrema detection techniques are visualized in Figure 4.31

H-L algorithms perform better using this neighborhood arrangement. Another approach using cube neighborhoods for all maxima detection produced a sparser collection of interest points that were less repeatable. This is a surprising result since the cube neighborhood is a tighter constraint. Table 4.3 shows the result of using the tightly constraining cube neighborhood on all three algorithms. DOG produces seven and 3.5

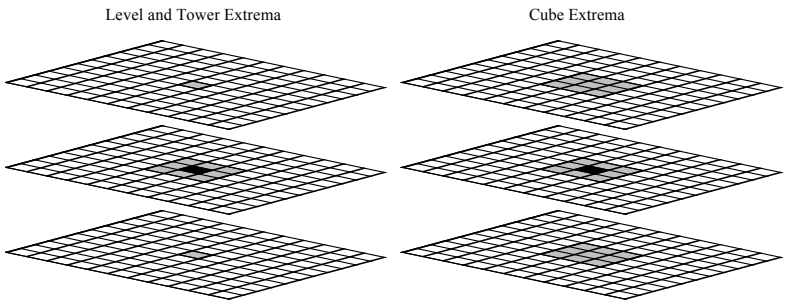

Figure 4.31: Two types of extrema detection. Level extrema are detected in the Harris or Hessian signal and tower extrema are detected in the Laplacian-of-Gaussian in [MS02]. Extrema are detected more rigorously in [Low03] using cube extrema. Use of cube extrema greatly reduces the number detected and negatively affects the repeatability of H-L interest points.

times as many interest points as Harris and Hessian-Laplace interest points, respectively. Even more interestingly, the repeatability of both H-L algorithms is negatively affected while accuracy is unaffected.

Table 4.3: Initial results with cube neighborhood extrema detection constraint for H-L algorithms. These data are produced from a different set of randomly selected source images. The ratio of interest point density is informative.

type	number	repeatability	accuracy
DOG/LOG	699,957	87 %	0.673
Harris-Laplacian	100,706	30 %	0.689
Hessian-Laplacian	199,337	67 %	0.691

These results lead us to implement the H-L interest point detectors according to their original authors. We suspect that the change in performance is primarily a function of density: blob and corner detectors are also edge detectors. Interest point repeatability from an edge detector increases proportionally with density, which may explain these findings.

These findings suggest that scale spaces must be constructed with consideration of the intended neighborhood function and provides a promising avenue for further research. They relate to the sinusoidal dependency of repeatability with scale in Sec-

tion 4.9. This work uses three levels per octave ($k = 2^{1/3}$) to construct a scale space as suggested by Lowe [Low03]. Mikolajczyk recommends $k = 1.2 \approx 2^{1/3}$ for the Harris-Laplace interest point detector. Such a small constant factor in the difference between levels of scale is unlikely to produce the behaviors in this section and Section 4.9. Further investigation is necessary, as both the method of extrema detection and the position of an interest point in scale space have a strong effect on repeatability.

Chapter 5

Conclusion

One million interest points are generated using three of the most popular selective attention algorithms - Lowe's DOG approach and two of Mikolajczyk and Schmid's Laplacian approaches: Harris-Laplace and Hessian-Laplace. The interest points are matched between randomly selected images from the CalTech-101 dataset and a target image that has undergone a random affine transformation. Matches are considered to be repeats when a distance and radius measure is satisfied.

We predict the repeatability of individual interest points by modeling them with generalized linear models. The models are produced by studying the 17 attributes of each interest point. No model is particularly informative, but each model suggests that there is an important relationship between an interest point's attributes and its repeatability.

The findings in the above experiments contribute to the following chapter. The previous experiments provide a framework for understanding how individual attributes affect repeatability. In this chapter, we perform short set of experiments combining this understanding to improve interest point repeatability. By combining the small effect of individual attributes on interest point repeatability we are able to produce a larger improvement. The conclusion has three components: recommendations to improve repeatability, a set of important observations that do not directly affect repeatability, and a number of promising avenues for future research based on our results. Section 5.1 pro-

vides our set of recommendations, based on either thresholding or interest point ranking, that can be used to improve the mean repeatability of interest points from any selective attention algorithm.

5.1 Summary experiments

This research has demonstrated that the repeatability of an interest point can be predicted from its bottom up attributes. A close examination of those attributes, however, reveals that they only slightly affect the probability that an interest point repeats. Our last experimental contribution combines the effect of these observations, improving repeatability by either thresholding or interest point ranking via multivariate regression.

5.1.1 Thresholds

The use of thresholding to improve repeatability and decrease density is supported by Lowe [Low03] and Mikolajczyk [MS02]. Lowe recommends discarding interest points where the ratio between Hessian eigenvalues $r > 10$ and when the determinant of the Hessian is negative. Mikolajczyk, though unsupported in this research, suggests discarding interest points with Harris corner score $R < 1000$. We extend on Lowe's thresholds by adding five additional thresholding decisions that improve repeatability performance. Seven thresholding decisions are supported in total, suggesting discarding interest points that do not fall inside of the following ranges:

- $hesdeterminant >= 0$
- $heslambda2 >= 0$
- $heslambda1/heslambda2 <= 5$
- $0.05 < \{xpos, ypos\} < 0.95$

- $harlambda1/harlambda2 <= 20$
- $entropy > 1$
- $dentropy < -1$

Implementing these seven decisions on our set of 10^6 interest points increases mean repeatability from 86% to 90.4% while reducing the total number of interest points to $174,291$. Thresholding in this fashion provides a fast and easy-to-implement solution to repeatability improvement. These thresholds can be applied only when applicable to the interest point detectors that depend on them or they can be computed for any interest point detector, improving the repeatability of that detector.

5.1.2 Multivariate generalized linear modeling

Having examined the individual contribution of each attribute on repeatability, we use a GLM to perform multiple linear regression. The next small set of experiments detail our results in selecting the best performing attributes according to AUC and $r_{E(Y),Y}$. The final experiments are separated into three groups: multiple-linear regression by attribute family, by the best fit attributes regardless of family, and by a hand-selected set of attributes whose logit functions share similar slopes. Finally Figures 5.1 and 5.2 shows the usable performance of three trained GLMs - one using a small subset of author selected attributes, one on the log normalized set of all attributes, and one on the log normalized set of Harris attributes.

Each attribute family contributes to interest point repeatability according to different theoretical foundations. We select the best attribute from each family, maximizing either AUC or $r_{E(Y),Y}$. Among original features we select $\{zpos, harlambda2, heslambda2, dentropy, value\}$, which maximize AUC, with results $AUC = 0.566, r_{E(Y),Y} = 0.1$. Maximizing $r_{E(Y),Y}$ we use $harlambda1$ instead of

$harlambda2$, improving both correlation and AUC to $r_{E(Y),Y} = 0.11$ and $AUC = 0.57$. Log normalization of attributes in this experiment offers no improvement.

The above experiment attempts to sample the best attribute from each family of features. Ignoring the family, we select the five attributes that contribute most to $r_{E(Y),Y}$: $heslambda2(0.07)$, $harlambda1(0.07)$, $dentropy(0.05)$, $zpos(0.03)$ and $entropy(0.03)$. The generalized model using these features improves AUC to 0.57 and $r_{E(Y),Y} = 0.12$ Again, log normalization of attributes in this experiment offers no improvement.

Finally, the authors look at the logit functions produced in each single-attribute experiment. A subset of attributes is selected that minimize the attribute-wise variation between the slope and curvature of the three logit functions. They are the most consistent across all three algorithms. We train a GLM using $zpos$, $harlambda1$, $log(harlambda2)$, $heslambda1$, $log(heslambda2)$ and $dentropy$ and further improved AUC to 0.585 and $r_{E(Y),Y} = 0.13$. This subset, referred to as the "consistency" fitting in Figure 5.1, scores nearly as well in AUC and $r_{E(Y),Y}$ as modeling the set of all 17 attributes. Figure 5.1 shows the performance of the two best performing generalized linear models: Our set of consistency-maximizing attributes and the multivariate GLM trained with all of the log-normalized data. Table 5.1 of GLM coefficients is included to assist in weighting interest points used by selective attention designers and implementers. The consistency GLM performs nearly as well as the GLM trained on all of the data. Training a GLM using the thresholded data from Section 5.1.1, unfortunately, offers no improvement. The two models perform nearly identically to the thresholding approach.

Of the three algorithms used for interest point detection in this work, only one benefits from analysis of its own attributes. Coefficients are included in Table 5.1 for a log normalized GLM (Figure 5.2) that can select the top percent of Harris interest points

Table 5.1: Log odds coefficients produced by a GLM trained to predict the repeatability of an interest point.

Attribute	LOG GLM	Small GLM	LOG(Harris) GLM
AUC	0.60	0.59	0.56
$r_{E(Y),Y}$	0.18	0.13	0.10
A priori	5.990	6.713	8.610
$xpos$	1.000		1.000
$ypos$	1.000		1.000
$zpos$	0.937	0.951	0.952
$log(harlambda1)$	0.933	1.000	0.981
$log(harlambda2)$	1.161	1.096	1.189
$log(hardeterminant)$	0.951		0.945
$log(heslambda1)$	1.187	1.003	
$log(heslambda2)$	1.092	0.970	
$log(hesdeterminant)$	0.876		
$entropy$	1.175		
$dentropy$	0.903	1.000	
$ddentropy$	1.071		
$log(value)$	1.104		
$log(truevalue)$	0.976		
$log(dx2)$	0.979		
$log(dy2)$	1.009		
$log(dz2)$	0.942		

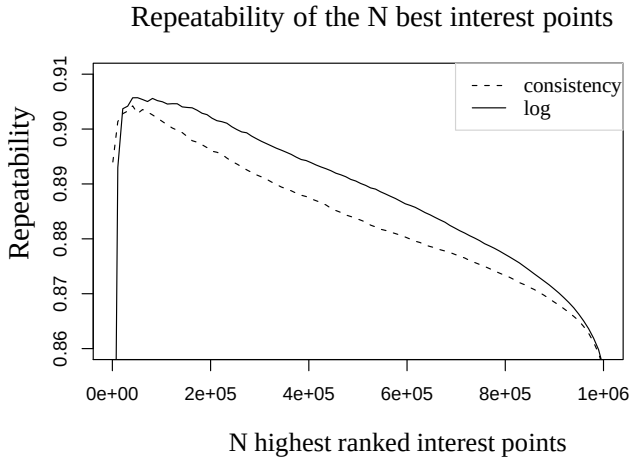

Figure 5.1: Fitting of a GLM fit to six author selected attributes and to a GLM fit to all 17 attributes including log normalized Harris, Hessian, and value families.

with 92.4% repeatability and the top tenth percentile with 90.6%. These coefficients should plug in to current implementations of Harris-Laplace (and we suspect Harris-Affine) with immediate repeatability improvement.

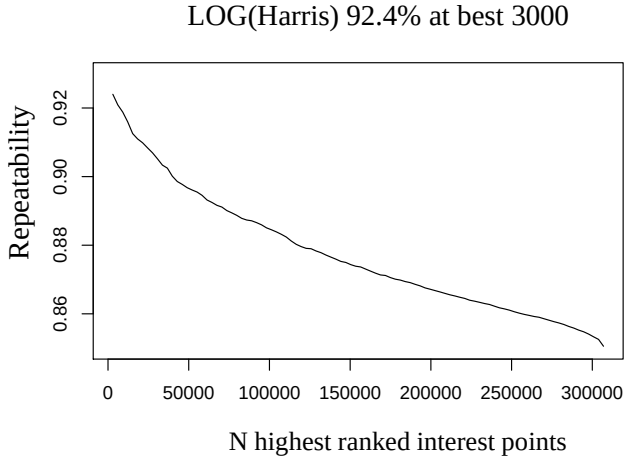

Figure 5.2: Fitting of a GLM fit to five author selected attributes and to a GLM fit to all 17 attributes including log normalized Harris, Hessian, and value families.

5.2 Discussion

We have seen that selection of a subset of highly repeatable interest points can be influenced by weighting interest points relative to their attributes. Achieving a set of extremely ($\approx 100\%$) repeatable interest points remains, however, a daunting task. Both thresholding and repeatability prediction increase the repeatability of interest points generated by three well-known interest point detectors by approximately 4%. Implementing our seven thresholds provides the same repeatability improvement while discarding more than four-fifths of the original set.

This work does not provide a means to select a small set of extremely repeatable interest points, but it has uncovered a number of areas where the behavior of scale spaces and interest point detectors is not fully understood. This section discusses some of those areas of opportunity.

In the close examination of individual attributes, DOG logit functions usually vary in slope significantly from H-L logit functions. This doesn't follow from one of the premises of this research – that interest points depend on similar attributes regardless of their generating algorithm. We believe this disparity is caused by their differing neighborhood detection method. More research will be necessary to properly describe the difference between these two techniques.

The magnitude of the Hessian second eigenvalue $heslambda2$ is more important to feature repeatability than its first eigenvalue. A large second eigenvalue depends on there being a large first eigenvalue, and as such any interest point with a large $heslambda2$ exhibits strong blob characteristics. The first eigenvalue is unimportant: interest points with a large first eigenvalue are edges in any case where the second eigenvalue is not similarly large. This relationship is inverted when considering the eigenvalues of the second moment matrix. Repeatable interest points rely more on a small $harlambda1$ than a large $harlambda2$. The ratio between these eigenvalues is another important

method for discarding edges from corner and blob selective attention algorithms.

The $value$ and $truevalue$ features are produced from the corresponding signal responsible for interest point detection in each algorithm, so the range of $value$ differs for each algorithm and should contribute little to generic interest point repeatability. We demonstrated in Section 4.5 that positively valued extrema are slightly more repeatable than negative extrema, which most likely produces this result. $truevalue$, being a sub-pixel optimized $value$, has a similar effect except on Harris-Laplace interest points. We believe this effect is caused by the lack of sub-pixel optimization nor smoothing of the derivatives that contribute to H.

Some vision applications depend on detection of a small number of highly repeatable interest points. Our thresholding guidelines and repeatability prediction can help to achieve this goal. An easier method for achieving this objective is by selecting only interest points with a large scale. As seen in Figure 4.9, repeatability decreases slightly as scale increases. Quantity, however, decreases logarithmically with scale. An application depending on a low quantity of highly repeatable interest points can then be selected from the higher levels of scale with little loss of repeatability. Selecting interest points with high scale is the easiest method to reduce quantity without losing repeatability. Repeatability can be improved cheaply up to 4% by discarding keypoints outside of the thresholds summarized above in Section 5.1.

5.3 Future work

Further investigation into the structure and theory of scale spaces is called for. There exist nearly unlimited parameterizations of scale space construction in the literature and their construction is not well standardized [BA83, ESM+06]. The repeatability of interest points fluctuate sinusoidally along each octave used in a scale space. In order to achieve true scale-invariance this effect needs to be minimized. Finally, the method of

extrema detection (Chapter 4.6) is strongly related to scale spaces and scale invariance and we believe a formal theory of its technique is welcome.

The $value$ family of features were surprisingly uninformative, given their description of neighborhood derivative information regardless of the generating algorithm. It is promising, then, to compute the local derivative information using the second moment matrix, which describes the structure of a local region. Computing neighborhood information using $D(\mathbf{x})$ is surprisingly uninformative because it does not take into account the direction of principal variance.

We believe that an opportunity exists to perform scene recognition by determining the affine structure of the entire image using the set of interest points. The affine shape of the probe image can be fit globally to images in the gallery with no transformation information. We suspect using this method there is a robust fit to the 8-dof correspondence problem without requiring the selection of four points of reference.

REFERENCES

[BA83] P. Burt and E. Adelson. The Laplacian Pyramid as a Compact Image Code. *Communications, IEEE Transactions on [legacy, pre-1988]*, 31(4):532–540, 1983.

[BL02] M. Brown and D.G. Lowe. Invariant features from interest point groups. *British Machine Vision Conference, Cardiff, Wales*, pages 656–665, 2002.

[BLGT06] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of SIFT features for face authentication. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 35. IEEE Computer Society Washington, DC, USA, 2006.

[CJ02] G. Carneiro and A.D. Jepson. Phase-Based Local Features. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 282–296, 2002.

[CRP02] J.L. Crowley, O. Riff, and J. Piater. Fast computation of characteristic scale using a half octave pyramid. In *Proceedings of the International Workshop on Cognitive Vision (CogVis 2002), Zurich, Switzerland*, 2002.

[dB93] J.M.H. du Buf. Responses of simple cells: events, interferences, and ambiguities. *Biological Cybernetics*, 68(4):321–333, 1993.

[DL04] B. Draper and A. Lionelle. Evaluation of selective attention under similarity transforms. In *Proc. of the Int'l Workshop on Attention and Performance in Computer Vision (WAPCV'03)*, pages 31–38, 2004.

[Dun98] J. Duncan. Converging levels of analysis in the cognitive neuroscience of visual attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373):1307, 1998.

[ECV] J. Eichhorn, O. Chapelle, and M. Vision. Object categorization with SVM: kernels for local features.

[ESM+06] R.S. Eaton, M.R. Stevens, J.C. McBride, G.T. Foil, and M.S. Snorrason. A Systems View of Scale Space. In *Computer Vision Systems, ICVS '06 IEEE International Conference*, 2006.

[FFFP07] L. Fei-Fei, R. Fergus, and P. Perona. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

[FPZ03] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2, 2003.

[Gil98] S. Gilles. *Robust Description and Matching of Images*. PhD thesis, PhD thesis, University of Oxford, 1998, 1998.

[GSKE+99] K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, and R. Malach. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1):187–203, 1999.

[HL03] J.S. Hare and P.H. Lewis. Scale Saliency: Applications in Visual Matching, Tracking and View-Based Object Recognition. *Proceedings, Distributed Multimedia Systems*, pages 436–440, 2003.

[HS88] C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, 15, 1988.

[IK01] Laurent Itti and Christof Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.

[IKN+98] L. Itti, C. Koch, E. Niebur, et al. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[Itt05] Laurent Itti. Models of Bottom-Up Attention and Saliency. *Neurobiology of Attention*, pages 576–582, 2005.

[JLK] Y.G. Jo, J.Y. Lee, and H. Kang. Segmentation Tracking and Recognition Based on Foreground-Background Absolute Features, Simplified SIFT, and Particle Filters. In *IEEE Congress on Evolutionary Computation, 2006. CEC 2006*, pages 1279–1284.

[JP87] JP Jones and LA Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[KB01] T. Kadir and M. Brady. Saliency, Scale and Image Description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[Koe84] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.

[KS] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2.

[KU85] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–227, 1985.

[KZB04] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. *European Conference on Computer Vision*, 1:228–241, 2004.

[Lin94] T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1):225–270, 1994.

[Lin98] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[Low99] D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, volume 2, pages 1150–1157. Kerkyra, Greece, 1999.

[Low03] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.

[LVKP02] F.F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14):9596, 2002.

[LW04] L. Ledwich and S. Williams. Reduced sift features for image retrieval and indoor localisation. *Australian Conf. on Robotics and Automation (ACRA)*, 2004.

[MBO07] AS Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 29(11):1927–1943, 2007.

[MCUP04] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[Mik02] K. Mikolajczyk. *Interest point detection invariant to affine transformations*. PhD thesis, PhD thesis, Institut National Polytechnique de Grenoble, 2002.

[MLS05] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. *Proc. ICCV*, 2:1792–1799, 2005.

[MP90] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 7(5):923–932, 1990.

[MP07] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 73(3):263–284, 2007.

[MS01] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Proc. ICCV*, volume 1, pages 525–531, 2001.

[MS02] K. Mikolajczyk and C. Schmid. An Affine Invariant Interest Point Detector. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 128–142, 2002.

[MS04] K. Mikolajczyk and C. Schmid. Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[MTS+05] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool. A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1):43–72, 2005.

[OFPK02] D.H. O'Connor, M.M. Fukui, M.A. Pinsk, and S. Kastner. Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5(11):1203–1209, 2002.

[OPFA06] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. Generic object recognition with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):416–431, 2006.

[OvWHM04] N. Ouerhani, R. von Wartburg, H. Hugli, and R. Muri. Empirical validation of the saliency-based model of visual attention. *Elec. Letters on Computer Vision and Image Analysis*, 3:13–24, 2004.

[PIIK05] R.J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. *Vision Research*, 45(18):2397–2416, 2005.

[PLN02] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123, 2002.

[SC00] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, 2000.

[SF03] Y. Sun and R. Fisher. Object-based visual attention for computer vision. *Artificial Intelligence*, 146(1):77–123, 2003.

[SH85] H. Spitzer and S. Hochstein. Simple-and complex-cell response dependences on stimulation parameters. *Journal of Neurophysiology*, 53(5):1244–1265, 1985.

[SI07] C. Siagian and L. Itti. Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE transactions on pattern analysis and machine intelligence*, 29(2):300, 2007.

[SL04] I. Skrypnyk and DG Lowe. Scene modelling, recognition and tracking with invariant image features. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality, 2004. ISMAR 2004*, pages 110–119, 2004.

[SM97] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[SMBI98] C. Schmid, R. Mohr, C. Bauckhage, and M. INRIA. Comparing and evaluating interest points. In *Computer Vision, 1998. Sixth International Conference on*, pages 230–235, 1998.

[TCW+95] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–545, 1995.

[TG80] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[TM08] Tinne Tuytelaars and Krystian Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Now Publishers Inc., Hanover, MA, USA, 2008.

[TYNT01] K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4:832–838, 2001.

[Wit87] A.P. Witkin. Scale-Space Filtering. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, 1987.

[WRKP04] D. Walther, U. Rutishauser, C. Koch, and P. Perona. On the usefulness of attention for object recognition. *Workshop on Attention and Performance in Computational Vision at ECCV*, pages 96–103, 2004.