HALF-SIFT: High-Accurate Localized Features for SIFT

Kai Cordes, Oliver Müller,
Bodo Rosenhahn, Jörn Ostermann

Institut für Informationsverarbeitung
Leibniz Universität Hannover

IEEE Workshop on
Feature Detectors and Descriptors: The State Of The Art and Beyond
CVPR, Miami, USA, 2009-06-21

Outline

Introduction

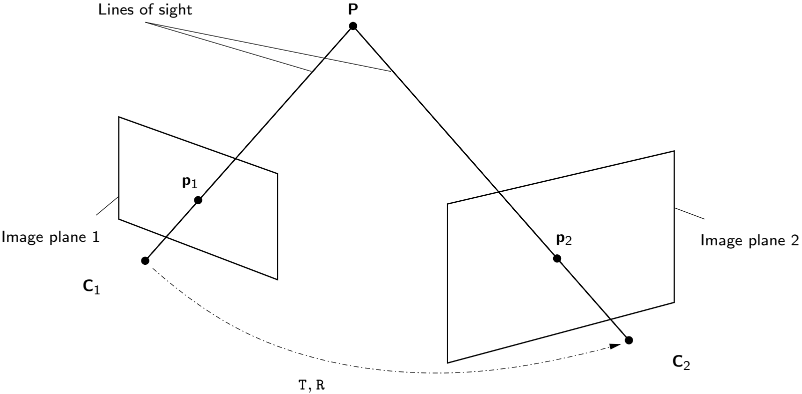
Localization of Features

Experimental Results

Conclusions

Introduction

- Detection of features in images requirement for computer vision



- Usage of SIFT framework
- Focus of our work: Precise detection of scale invariant feature points

Workflow of Scale Invariant Feature Transform

Image

↓

Detection of Scale-Space Extrema

↓

Localization of Features

↓

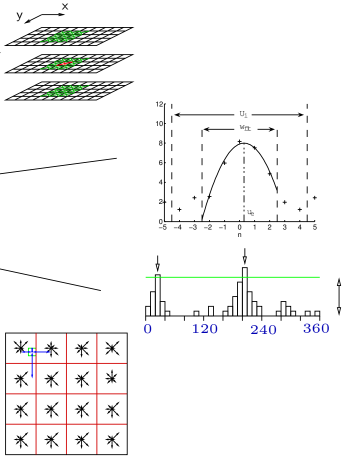
Orientation Assignment

↓

Descriptor Calculation

↓

Image Features



Localization: Subpel and Subscale Interpolation

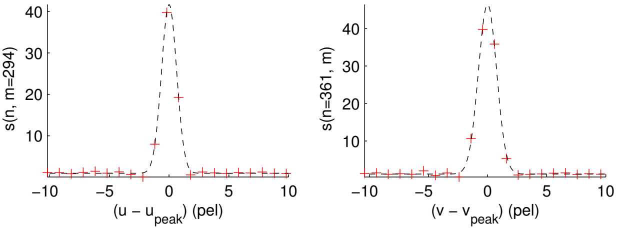
- SIFT: Parabolic interpolation *(Taylor)*

$$D(\mathbf{x}) = D(\mathbf{x}_0) + \frac{\partial D(\mathbf{x}_0)^\top}{\partial \mathbf{x}}\mathbf{x}^\top + \frac{1}{2}\mathbf{x}^\top \frac{\partial^2 D(\mathbf{x}_0)}{\partial \mathbf{x}^2}\mathbf{x} \quad (1)$$

  - $\mathbf{x} = (x, y, s)$ subpel feature localization
  - $D(\mathbf{x})$: approximation of the Difference of Gaussian (DoG)
  - $D(\mathbf{x}_0)$: DoG at fullpel sample point (x_0, y_0, s_0)
- Solution with *Hessian Matrix*

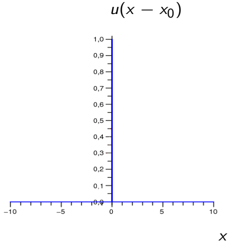
Motivation

- Image signal characteristics are not parabolic
- Assumption: Impulse response of lens-aperture system is Gaussian
  - Impulse response to real camera (Canon XL-1):



- Main question: Proper approximation function for SIFT keypoints

Motivation



Feature Point

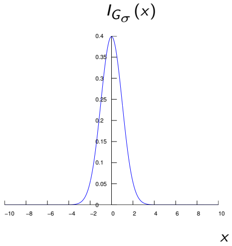
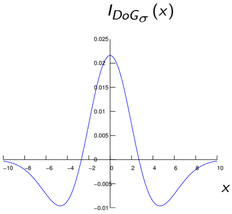


Image Signal

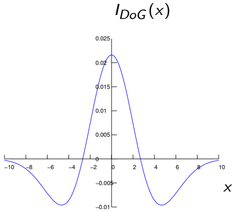


DoG signal curvature

feature detection at scale s

result

HALF-SIFT: DoG Signal Approximation Proposal



DoG signal curvature

- Difference of Gaussians Regression
  - Multiple Extrema
  - More challenging optimization
- Gaussian Regression

Regression Function Proposal (1)

- Gaussian Regression Function:

$$G_{\mathbf{x}_G,\Sigma}(\mathbf{x}) = \frac{r_G}{\sqrt{|\Sigma|}} \cdot e^{-\frac{1}{2}((\mathbf{x}-\mathbf{x}_G)^\top \Sigma^{-1}(\mathbf{x}-\mathbf{x}_G))} \tag{2}$$

- Covariance matrix $\Sigma = \begin{pmatrix} a^2 & b \\ b & c^2 \end{pmatrix} \Rightarrow$ rotated, scaled ellipse
- $\mathbf{x}_G = (x_0, y_0)$ subpel position
- Parameter vector $\mathbf{p_G} = (x_0, y_0, a, b, c, r_G)$
- Levenberg-Marquardt (LM) optimization

Regression Function Proposal (2)

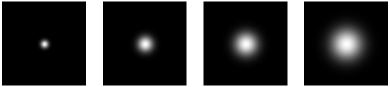
- Difference of Gaussians (DoG) Regression Function:

$$\begin{aligned} D_{\mathbf{x}_D,\sigma} &= r_D(G_{\mathbf{x}_D,\Sigma_\sigma} - G_{\mathbf{x}_D,\Sigma_{k\sigma}}) * G_{\mathbf{x}_D,\Sigma} \\ &= r_D(G_{\mathbf{x}_D,\Sigma_\sigma+\Sigma} - G_{\mathbf{x}_D,\Sigma_{k\sigma}+\Sigma}) \end{aligned} \tag{3}$$

- Known distance k between scales
- Parameter vector $\mathbf{p_D} = (x_0, y_0, a, b, c, r_D)$
- Levenberg-Marquardt (LM) optimization

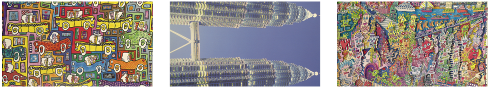
Experimental Setups

1. Synthetically constructed Gaussian blob images (64x64):

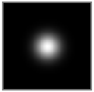


2. Image pairs[1], estimated homographies (800x640, 1000x700)

3. Image pairs, constant extrinsics, varying illumination (1920x1080)

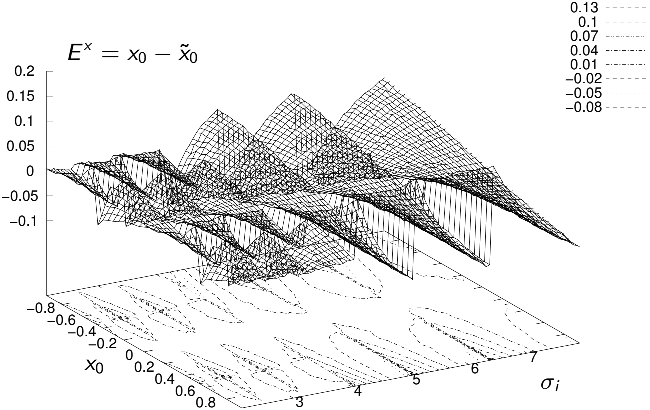


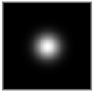
[1] www.robots.ox.ac.uk/~vgg/research/affine/index.html



Gaussian blobs

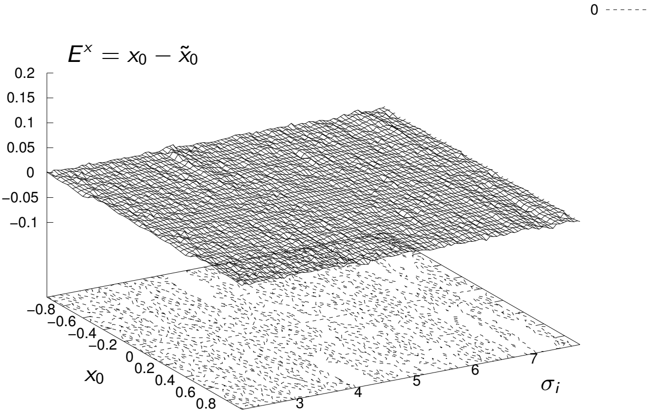
Error: standard SIFT localization x_0



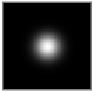


Gaussian blobs

Error: proposed HALF-SIFT DoG localization x_0



Gaussian blobs



Maximum error of SIFT and proposed DoG

i	σ_i	x_0	$E^{max\_x}_{\mathbf{PARAB}}$	$E^{max\_x}_{\mathbf{GAUSS}}$	$E^{max\_x}_{\mathbf{DOG}}$
1..3	1.6..3.9	$-0.5..0.5$	0.072	0.006	0.006
4..6	4.0..7.9	$-1.0..1.0$	0.142	0.009	0.009
7..8	8.0..9.9	$-2.0..2.0$	0.297	0.014	0.021

Natural image pairs, given homography

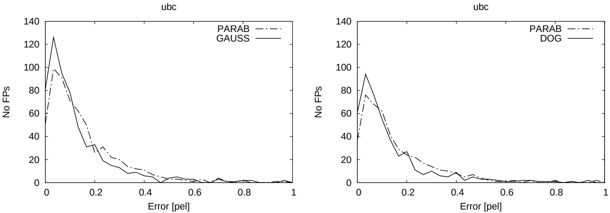
- Repeatability criterion to verify corresponding feature points

$$|\mathbf{x}_a - H \cdot \mathbf{x}_b| < d_{thres}$$

- Only correspondences detected with both methods
- Feature point error classification into 30 bins

Natural image pairs

Error bins of *ubc 1-2* (homography $\mathtt{H} = \mathtt{E}$)



ubc Pair	Mean Error PARAB	Mean Error GAUSS	Mean Error PARAB	Mean Error DOG
1-2	0.17	0.15	0.17	0.15
1-3	0.27	0.25	0.26	0.24
1-4	0.35	0.33	0.36	0.34

Natural image pairs



Estimated homographies *graf*, *bikes*, ...

- Lead to mapping error > 0.4 pel per feature
- No reasonable ground truth homography

Pairs with exact homography ($\mathtt{H} = \mathtt{E}$, *ubc*)

- Accuracy gain of up to 13.9% for Gaussian
- Accuracy gain of up to 13.6% for DoG

Image Pairs with Illumination Changes

Three scenarios with constant extrinsics

1. Intrinsics constant $\Rightarrow$ 2^{ND} image lower dynamics

2. *Autoiris* $\Rightarrow$ compensation with aperture of the camera

3. *Automatic gain* $\Rightarrow$ compensation with amplifier of the camera

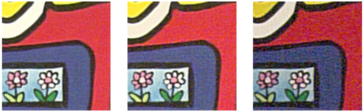
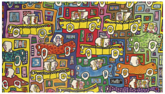
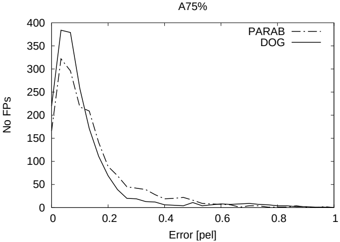
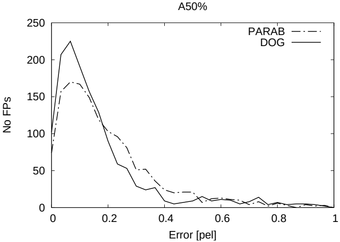


Image Pairs with Illumination Changes



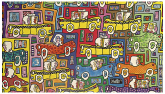
DoG error bins of *cars* $A^{75\%}$ and $A^{50\%}$



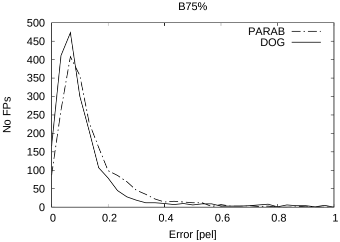


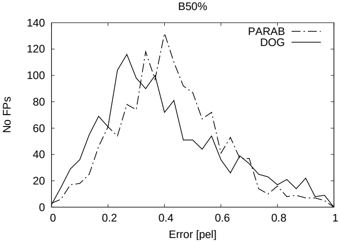
	Mean Error		Mean Error	
scene	**PARAB**	**GAUSS**	**PARAB**	**DOG**
$A^{75\%}$	0.15	0.13	0.16	0.13
$A^{50\%}$	0.21	0.18	0.21	0.19

Image Pairs with Illumination Changes



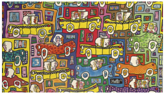
DoG error bins of *cars* $B^{75\%}$ and $B^{50\%}$



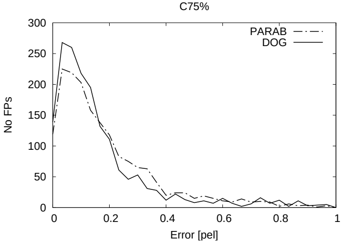


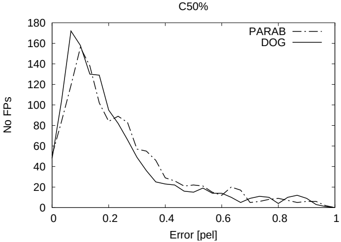
scene	Mean Error **PARAB**	Mean Error **GAUSS**	Mean Error **PARAB**	Mean Error **DOG**
$B^{75\%}$	0.15	0.14	0.15	0.13
$B^{50\%}$	0.47	0.46	0.44	0.41

Image Pairs with Illumination Changes



DoG error bins of *cars* $C^{75\%}$ and $C^{50\%}$



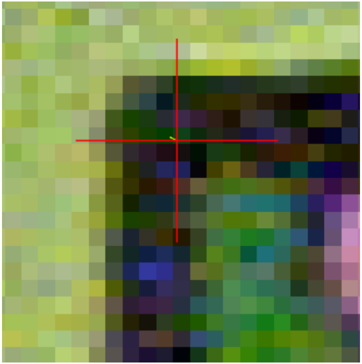


	Mean Error		Mean Error	
scene	**PARAB**	**GAUSS**	**PARAB**	**DOG**
$C^{75\%}$	0.20	0.18	0.20	0.18
$C^{50\%}$	0.26	0.24	0.26	0.24

Image Pairs with Illumination Changes

- Accuracy gain of up to 11.0% for Gaussian
- Accuracy gain of up to 15.6% for DoG

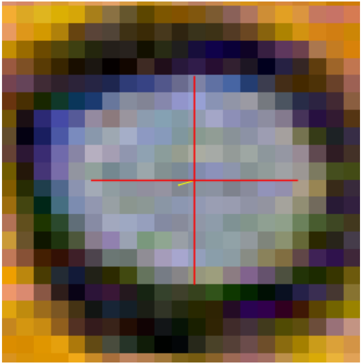
$E_{PARA} = 0.6604$



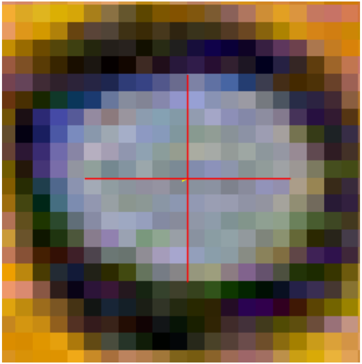
$E_{DOG} = 0.4159$

Image Pairs with Illumination Changes

- Accuracy gain of up to 11.0% for Gaussian
- Accuracy gain of up to 15.6% for DoG



$E_{PARA} = 0.9368$



$E_{DOG} = 0.3311$

Summary

- Keypoint localization improved using Gaussian / DoG regression instead of parabolic interpolation (standard SIFT)
- Functionality validated with synthetic data
- Natural image pairs show accuracy gain of up to
  - 15.6% for DoG and
  - 13.9% for Gaussian
- More results
  - `http://www.tnt.uni-hannover.de/~cordes/half-sift/`