An Analysis of Scale-space Sampling in SIFT

Ives Rey-Otero[†] Jean-Michel Morel[†] Mauricio Delbracio [⋆†]

[†]CMLA, ENS-Cachan, France [⋆]ECE, Duke University, USA

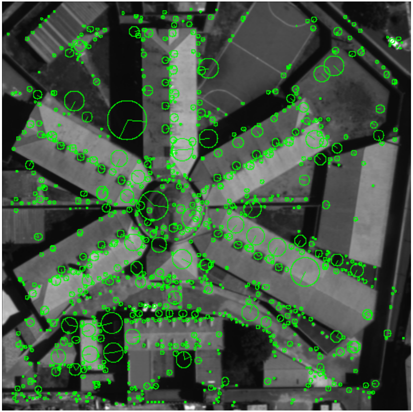
IEEE ICIP 2014, Paris

Overview

In this article, we propose:

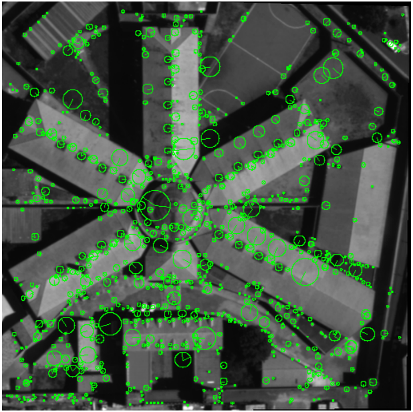
- a study of SIFT
- an experimental framework consistent with SIFT's camera model
- an analysis of detection stability and invariance
- a study on the influence of
  - scale-space sampling
  - image aliasing
  - thresholds aiming at discarding unstable detections

Overview



Detection stability ?

Overview



Detection stability ?

Overview

This paper is not:

a variant of SIFT

- SURF: Speeded Up Robust Feature (Bay et al. 2006)
- Affine SIFT (Yu and Morel, 2009)
- Spectral-SIFT (Koutaki and Kumamoto, 2014)

a new feature descriptor

- On affine invariant descriptors related to SIFT (Sadek and Caselles, 2012)
- BRIEF (Calonder et al. 2010)
- K-means Hashing (He et al. 2013)

a benchmark

- A comparison of affine region detectors (Mikolajczyk et al. 2005)

Overview

Study the influence of *scale-space sampling* on SIFT

SIFT applied with 3 different sampling settings

SIFT overview

The detection step:

- Assumes that the input image $u(\mathbf{x})$ has a Gaussian camera blur of c
- Gaussian scale-space $v(\sigma, \mathbf{x}) = G_{\sqrt{\sigma^2 - c^2}} u(\mathbf{x})$
- Differential operator: DoG (difference of Gaussians)
- Extract discrete 3D extrema
- Refine position of 3D extrema (local quadratic model)
- Filter unstable keypoints (thresholds)

SIFT camera model

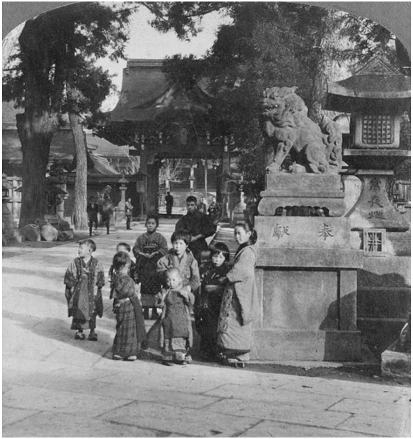
The camera model adopted by SIFT approximates the point spread function (PSF) by a *Gaussian kernel of standard deviation c*.

$$\mathbf{u} =: \mathbf{S}_1 G_c H \mathcal{T} R u_0$$

- $\mathbf{S}_1$ sampling operator
- Gaussian kernel of standard deviation c
- H homothety
- $\mathcal{T}$ translation
- R rotation

SIFT overview

The camera model adopted by SIFT approximates the point spread function (PSF) by a Gaussian kernel of standard deviation c.

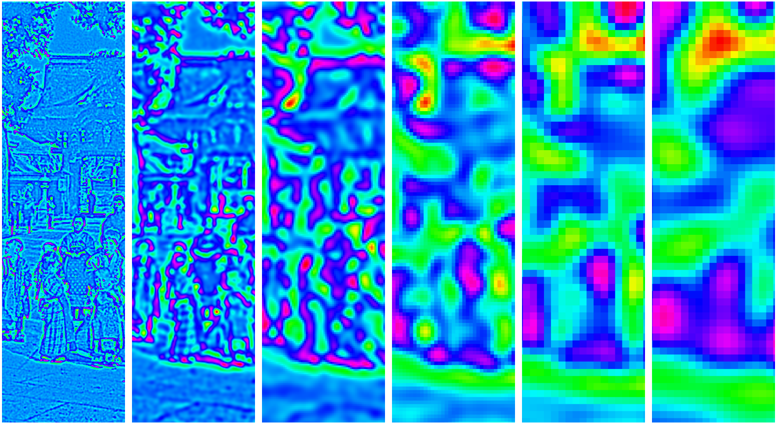


SIFT overview

Compute the Gaussian scale-space $v(\sigma, \mathbf{x}) = G_{\sqrt{\sigma^2 - c^2}} u(\mathbf{x})$

SIFT overview

Compute differential operator: DoG (difference of Gaussians)



Extract 3D extrema

Theoretically invariant to zoom outs

Let $\mathbf{u}_\lambda$ and $\mathbf{u}_\mu$ be two different acquisitions at two different distances

$$\mathbf{u}_\lambda = \mathbf{S}_1 G_c H_\lambda u_0 \qquad \mathbf{u}_\mu = \mathbf{S}_1 G_c H_\mu u_0$$

Their respective scale-space

$$v_\lambda(\sigma, \mathbf{x}) = G_\sigma H_\lambda u_0(\mathbf{x}) \qquad v_\mu(\sigma, \mathbf{x}) = G_\sigma H_\mu u_0(\mathbf{x})$$

Reparameterizations of $v_0(\sigma, \mathbf{x}) = G_\sigma u_0(\sigma, \mathbf{x})$

$$v_\mu(\sigma/\mu, \mathbf{x}/\mu) = v_\mu(\sigma/\lambda, \mathbf{x}/\lambda)$$

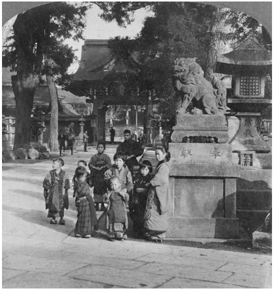
Does this perfect invariance hold in practice ?

The architecture of *digital* Gaussian scale-space

A set of digital images with various level of blur σ and sampled at various rates δ.

The architecture of *digital* Gaussian scale-space

Supersample by a factor $1/\delta_{\min}$ (default $1/\delta_{\min} = 2$)
Add extra blur $(G_{(\sigma_{\min}{}^2 - c^2)^{1/2}/\delta_{\min}})$ to reach the minimal level of blur $\sigma_{\min}$ (default $\sigma_{\min} = 0.8$)



blur c

supersampling factor $1/\delta_{\min} > 1$
blur $\sigma_{\min} > c$

The architecture of *digital* Gaussian scale-space

The scale-space is split into octaves, subsets of n_{spo} images (default $n_{\text{spo}} = 3$) sharing the same sampling rate $1/\delta$

Blurs follow a geometric progression

$$\sigma = \sigma_{\text{min}} 2^{s/n_{\text{spo}}}$$

Increase the scale-space sampling rates:

- $\nearrow 1/\delta_{\text{min}}$
- $\nearrow n_{\text{spo}}$

Simulating the camera model

How to simulate snapshots having a given level of blur c

- Take a large image u_{in} with unknown level of blur c_{in}
- Apply a Gaussian filtering of standard deviation $s \times c$
- Apply a subsampling of factor $s >> 1$

$$u_{\text{simul}} = S_{1/s} G_{s \times c} u_{\text{in}}$$

The approximated level of blur is

$$\sqrt{c^2 + (c_{\text{in}}/s)^2} \approx c$$

The experimental setup

Measure the invariance level by accurately simulating image pairs related through a scale change, a translation or a blur

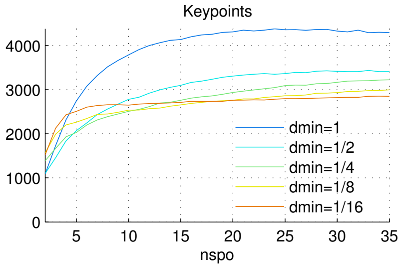
Comparing the set of detected keypoints

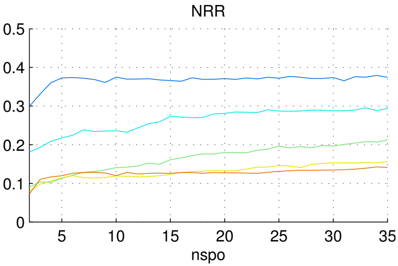
The *non repeatability ratio* (NRR) is the number of keypoints detected in one image but not detected on the other at its expected position ($+/-0.25px$ in space, $+/-2^{1/4}s$ relatively in scale) divided by the total number of detected keypoints

Influence of scale-space sampling - Translation

Subpixel translation of 0.25px

Compare the sets of keypoints for various scale-space settings

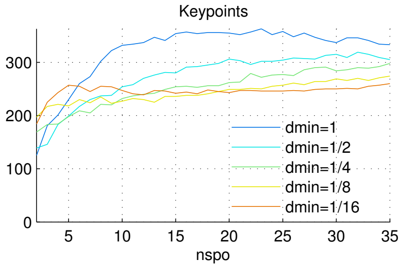


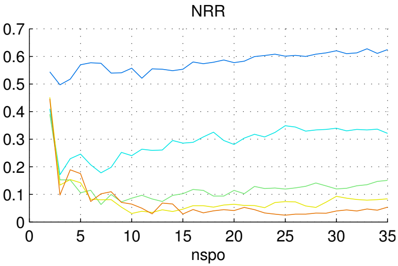


Influence of scale-space sampling - Zoom-out

$2.15\times$ zoom-out

Compare the sets of keypoints for various scale-space settings

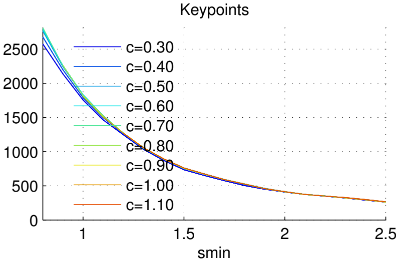


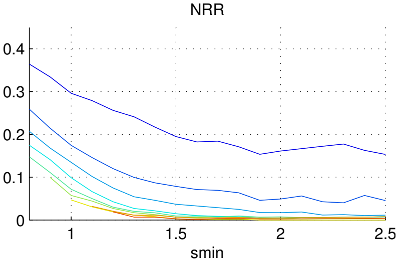


Influence of image blur

Stability varying the level of blur c in the input image
$(0.30 \leq c \leq 1.10$, no image aliasing for $c > 0.75)$

Subpixel translation of 0.25px
Keypoint stability to the image blur and the detection scale

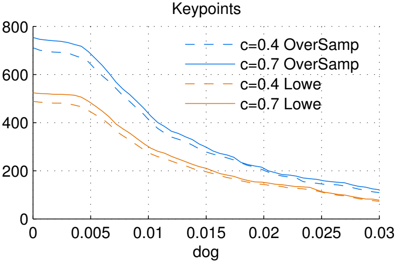


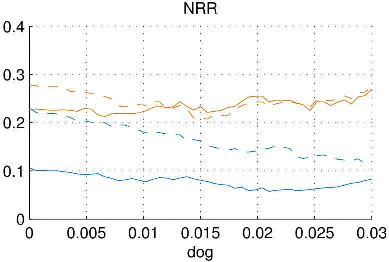


The DoG threshold

Is the DoG threshold efficient at *discarding unstable keypoints* ?

Subpixel translation of 0.25*px*
Increasing the DoG threshold





The DoG threshold fails to significantly improve the overall stability of keypoints.

Conclusions

- Invariance is limited by insufficient sampling of the Gaussian scale-space
- Invariance is limited by image aliasing
- The DoG threshold is not efficient

Future work:

- Extend this analysis in the case where the input image blur is not consistent with SIFT's camera model
- Analyse the influence of image noise