Heavy Tailed Distances for Gradient Based Image Descriptors

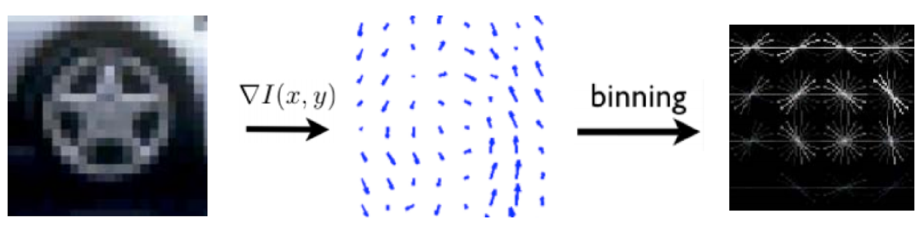
Yangqing Jia and Trevor Darrell
UC Berkeley EECS {jiayq,trevor}@eecs.berkeley.edu

0. Summary

- Most computer vision algorithms use the Euclidean distance when comparing image descriptors against each other, derived from a Gaussian noise assumption.
- We examined why this may be so, and showed the heavy-tailed nature of the noise in gradient-based image descriptors.
- We proposed a heavy-tailed distance measure derived from heavy-tailed distributions that empirically works better than existing distance measures.

1. Image Descriptors

- In the recent years the statistics of oriented gradients have been shown to form particularly effective image representations.
- Examples: SIFT (Lowe 2004), HOG (Dalal 2005), etc.

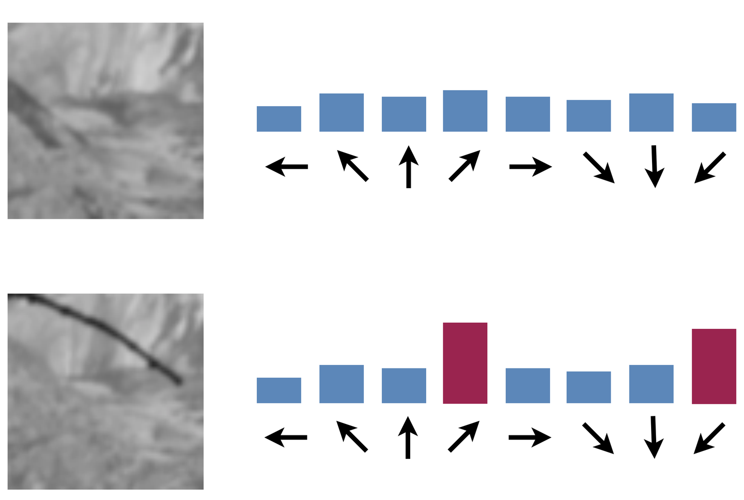


(Image Courtesy of Fritz et al. NIPS'09)

- Applications: local image descriptor matching (Snavely 2006), image classification (Yang 2009), object detection (Dalal 2005), etc.
- Biological evidences include orientation selectivity (Hubel 1967) and Gabor-like filters in V1 (Olshausen 1997).

2. Image Statistics

- Matching Image Descriptors shows a heavy-tailed noise (see right).
- Generative explanation: natural disturbance usually results sparse changes in the oriented histograms, with potentially high magnitudes.



- Heavy-tailed properties have been observed in natural image statistics, optical flow, stereo vision, shape from shading, etc.

References

[1] S Winder and M Brown. Learning local image descriptors. In *CVPR*, 2007.

[2] P Chen, Y Chen, and M Rao. Metrics defined by Bregman divergences. CMS, 2008.

3. The Distribution

- Given a prototype μ, we want to find $p(x|\mu)$.
- A common approach to cope with heavy tails is to use L_1 distance, corresponding to the Laplace distribution:

$$p(x|\mu;\lambda) = \frac{\lambda}{2}\exp(-\lambda|x-\mu|)$$

The tail is still exponentially bounded.

- We adopt the hierarchical Bayesian idea by introducing a Gamma prior over λ:

$$p(\lambda) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha-1}\beta^{\alpha}e^{-\beta\lambda}\mathrm{d}\lambda$$

This yields the Gamma-compound-Laplace distribution

$$p(x|\mu;\alpha,\beta) = \frac{1}{2}\alpha\beta^{\alpha}(|x-\mu|+\beta)^{-\alpha-1}$$

- The GCL distribution is heavy-tailed.

4. Hypothesis Test

- The hypothesis test is widely adopted to test if certain statistical models fit observations.
- The likelihood ratio test with data points x and y:
  - Null hypothesis: same prototype μ_{xy}
  - General: different prototypes μ_x, μ_y
- Conventionally the test is used to *accept* or *reject* the null hypothesis.
- Score:

$$s(x,y) = \frac{p(x|\hat{\mu}_{xy})p(y|\hat{\mu}_{xy})}{p(x|\hat{\mu}_x)p(y|\hat{\mu}_y)}$$

- ML estimations μ_{xy} for the GCL distribution is closed-form.
- We define the *likelihood ratio distance* as

$$d(x,y) = \sqrt{-\log(s(x,y))}$$

(Note: not a metric for arbitrary $p(x|\mu)$)

5. Heavy-tailed Distance Measure

Theorem: If the distribution $p(x|\mu)$ can be written as $p(x|\mu) = \exp(-f(x-\mu))b(x)$, where $f(t)$ is a non-constant quasi-sconvex function w.r.t. t that satisfies $f''(t) \leq 0, \forall t \in \mathbb{R}\backslash\{0\}$, then the distance defined above is a metric.

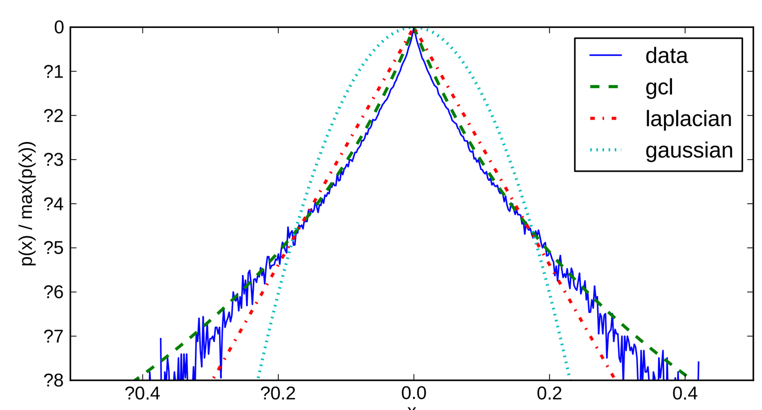
Relationship to Existing Distances (See [2] for discussion on non-regular exp family)

Distribution	Gaussian	Laplace	Multinomial	Regular Exp Family
Distance	Euclidean	L_1	Jensen-Shannon	Jensen-Bregman
$d^2(x,y)$	$\|x-y\|_2^2$	$\|x-y\|_1$	$(D_B(x\|\mu_{xy}) + D_B(y\|\mu_{xy}))/2$	

Distance for the GCL distribution:

$$d^2(x,y) = (\alpha+1)(\log(|x-y|+\beta) - \log\beta)$$

6. Does It Fit?



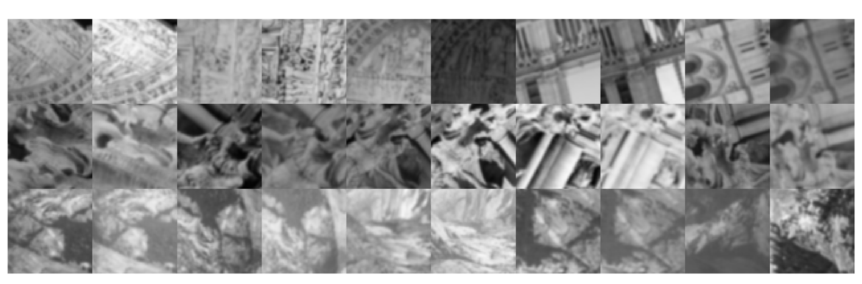
- The GCL distribution fits the highly kurtotic noise distribution better than baselines.
- Parameter Estimation:

$$\alpha \leftarrow n\left(\sum_{i=1}^{n}\log(|x_i|+\beta) - n\log(\beta)\right)^{-1}, \beta \leftarrow \beta - \frac{l'(\beta)}{l''(\beta)}$$

$$l'(\beta) = \frac{n\alpha}{\beta} - \sum_{i=1}^{n}\frac{\alpha+1}{|x_i|+\beta}, l''(\beta) = \sum_{i=1}^{n}\frac{\alpha+1}{(|x_i|+\beta)^2} - \frac{n\alpha}{\beta^2}$$

7. Experiments and Future Work

- We used the Photo tourism data [1] containing matching local image patches.
- Feature: SIFT (on patches with jittering effect.

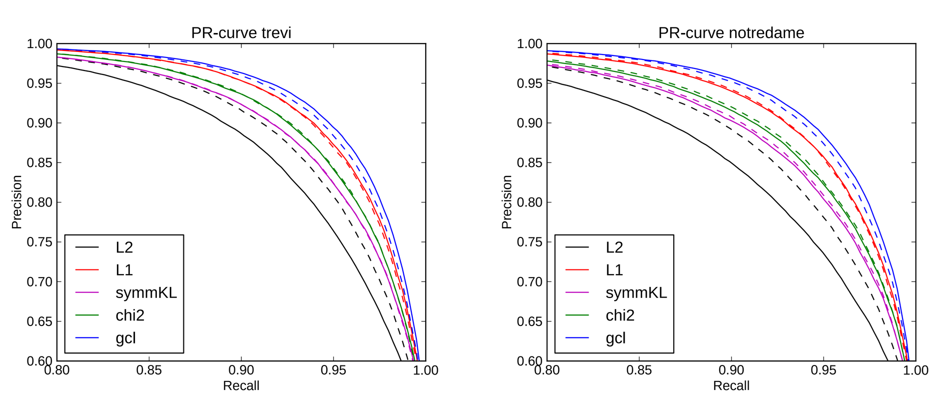


- Protocol: Predicting matches based on pairwise distance.

Table: Average precision on Halfdome w/ and w/o SIFT feature thresholding

d	L_2	L_1	χ^2	GCL
w/o	94.51	96.75	95.42	**98.19**
w/	95.55	96.90	95.64	97.21

Figure: Precision-Recall curves



Future Direction:

- Learning image descriptors and distance measures with heavy-tailed properties.
- Latent variable models (such as PCA and sparse coding) with heavy-tailed noise assumption.
- Discriminative models for image classification.