# Heavy-tailed Distance for Local Image Descriptors Supplementary Materials

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

## 1 Proof of Theorem 3.1

Theorem 3.1 in the paper states that:

**Theorem 3.1.** 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-convex function w.r.t. $t$ that satisfies $f''(t) \leq 0, \forall t \in \mathbb{R}\backslash\{0\}$, then the distance defined as

$$d(x,y) = \sqrt{-\log\left(\frac{p(x|\hat{\mu}_{xy})p(y|\hat{\mu}_{xy})}{p(x|\hat{\mu}_x)p(y|\hat{\mu}_y)}\right)} \tag{1}$$

is a metric.

Before proving the theorem, we propose the following lemmas:

**Lemma 1.1.** *If a function $d(x,y)$ defined on $\mathbb{R}\times\mathbb{R}\rightarrow\mathbb{R}$ is a distance metric, then $\sqrt{d(x,y)}$ is also a distance metric.*

*Proof.* Triangle inequality is the only nontrivial part. It is easy to see that $\forall x,y,z\in\mathbb{R}$,

$$\begin{aligned}\left(\sqrt{d(x,y)}+\sqrt{d(y,z)}\right)^2 &= d(x,y)+d(y,z)+2\sqrt{d(x,y)d(y,z)}\\ &\geq d(x,z)++2\sqrt{d(x,y)d(y,z)}\\ &\geq d(x,z)\end{aligned}$$

□

**Lemma 1.2.** *If function $f(t)$ is defined as in Theorem 3.1, then we have:*
*(1) the minimizer $\hat{t} = \arg\min_t f(t)$ is $0$, and*
*(2) the minimizer $\hat{\mu}_{xy} = \arg\min_\mu f(x-\mu)+f(y-\mu)$ is either $x$ or $y$.*
*(3) the function $g(t) = \min(f(t), f(-t)) - f(0)$ is monotonically increasing and concave in $\mathbb{R}^+\cup\{0\}$, and $g(0)=0$.*

*Proof.* The first result about $\mu_x$ is the direct result of condition (b). For the second result, assuming that $\mu_{xy}$ is neither $x$ or $y$, because $f(x)$ is smooth on $\mathbb{R}\backslash\{0\}$,we have

$$\frac{\partial f(x-\mu)}{\partial\mu}+\frac{\partial f(y-\mu)}{\partial\mu}=0, \frac{\partial^2 f(x-\mu)}{\partial\mu^2}+\frac{\partial^2 f(y-\mu)}{\partial\mu^2}>0 \tag{2}$$

the result of the second-order derivative contradicts condition (b). Thus we have $\mu_{xy}=x$ or $\mu_{xy}=y$. The third result comes from the fact that the minimum of two concave functions is still a concave function. □

We then prove Theorem 3.1.

*Proof.* (**Proof of Theorem 3.1**) We write $d^2(x,y)$ as

$$d^2(x,y) = f(x-\mu_{xy}) + f(y-\mu_{xy}) - f(x-\mu_x) - f(y-\mu_y). \tag{3}$$

According to Lemma 1.2, this is equivalent to $d^2(x,y) = g(|x-y|)$. Given 3 arbitrary real numbers $x, y$, and $z$, without loss of generality, we assume that they are ordered: $x < y < z$, and define $a = y - x, b = z - y$. There are three possible triangle inequalities to prove:

$$\begin{aligned} g(a) + g(b) &\geq g(a+b) && (4) \\ g(a+b) + g(a) &\geq g(b) && (5) \\ g(a+b) + g(b) &\geq g(a) && (6) \\ & && (7) \end{aligned}$$

The second and third inequalities are straightforward from the monotonicity of $g$. The first inequality holds due to the subadditivity of $g(\cdot)$. Thus $d^2(x,y)$ is a metric, and by Lemma 1.1, $d(x,y)$ is also a metric.

We note here that $f''(t) \leq 0$ may not be a necessity condition, as what we need is the concavity of $min(f(t), f(-t))$. However, for many heavy-tailed distributions (such as the GCL distribution), $f(t)$ is symmetric with respect to the origin point, and the concavity of $min(f(t), f(-t))$ is equivalent to requiring $f''(t) \leq 0$. □

## 2 Large Figures

Figure 1 is a larger version of Figure 4 in the paper for better readability.

## 3 Additional Experimental Results

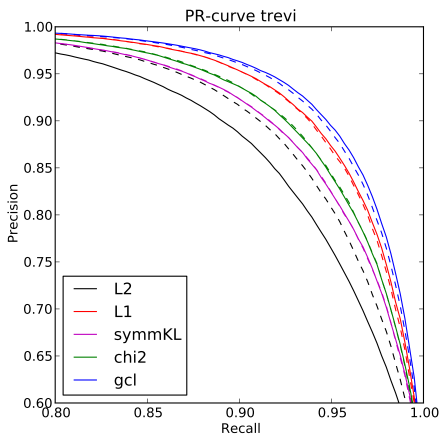
We present the pr curves, average precision and 99% false positive rate of the experiments carried out on the original Photo Tourism data without jitter effects. Figure 2 shows the curves and Table 1 summarizes the numbers. Note that, without noise, all the distance metrics perform well. This is due to the way the data is collected, as stated in the original paper [1]:

> "Since our evaluation data sets were obtained by projecting 3D points into 2D images we expected that the normalization for scale and orientation and the accuracy of spatial correspondence of our patches would be much greater than that obtained directly from raw interest point detections."

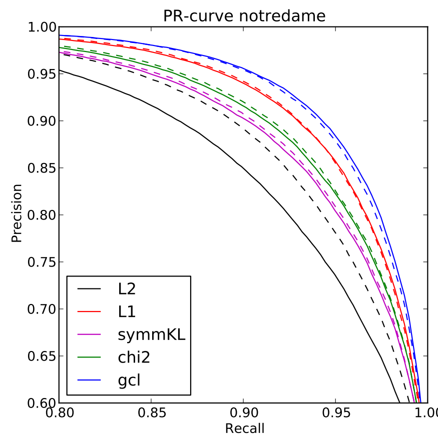
This is why both [1] and our paper introduce jitter effects into the experiments. The no-noise case is presented here for the sake of completeness. We note that although the performance of base distance measures is already high, our heavy-tailed distance still yields noticeable improvement over the baselines, showing the advantage of taking into the underlying heavy-tailed distribution into consideration.

## References

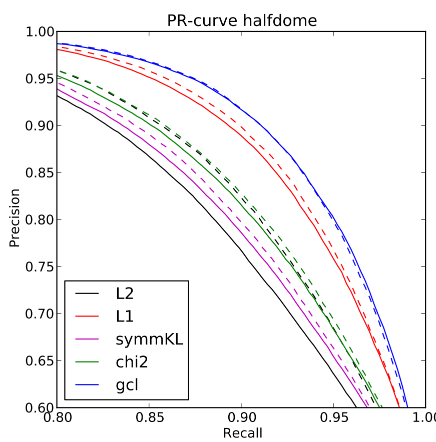
[1] S Winder and M Brown. Learning local image descriptors. In *CVPR*, 2007.




(a) trevi

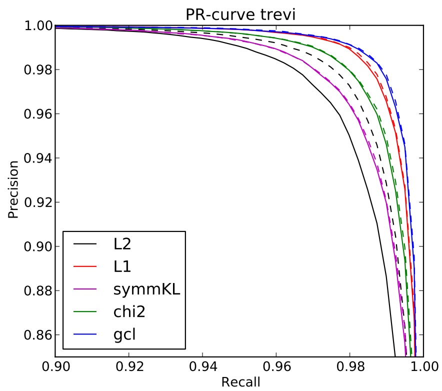



(b) notredame

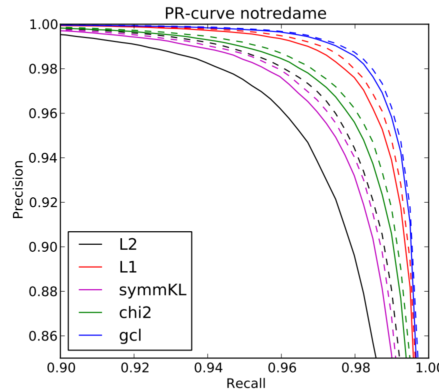



(c) halfdome

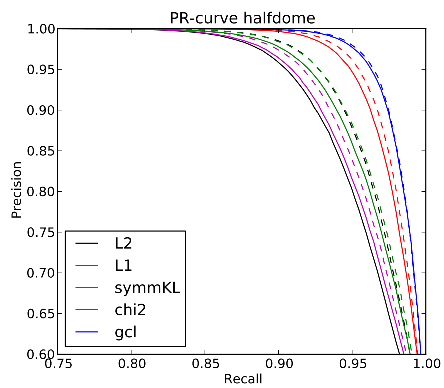
Figure 1: The mean precision-recall curve over 20 independent runs. In the figure, solid lines are experiments using features that are l2 normalized, and dashed lines using features thresholded and re-normalized. Best viewed in color.




(a) trevi




(b) notredame




(c) halfdome

Figure 2: The mean precision-recall curve over 20 independent runs, on the features extracted from raw patches without jitter effects. In the figure, solid lines are experiments using features that are l2 normalized, and dashed lines using features thresholded and re-normalized. Best viewed in color. Note that axis scales may differ between figures.

| AP | $L_2$ | $L_1$ | SymmKL | $\chi^2$ | GCL |
|---|---|---|---|---|---|
| trevi-l2 | 99.61±0.04 | 99.85±0.02 | 99.70±0.03 | 99.79±0.03 | **99.88±0.02** |
| trevi-thres | 99.73±0.03 | 99.86±0.02 | 99.71±0.03 | 99.80±0.03 | **99.89±0.02** |
| notre-l2 | 99.31±0.05 | 99.76±0.02 | 99.48±0.05 | 99.64±0.03 | **99.82±0.02** |
| notre-thres | 99.57±0.03 | 99.79±0.02 | 99.53±0.04 | 99.68±0.03 | **99.83±0.02** |
| halfd-l2 | 97.61±0.11 | 98.84±0.07 | 97.76±0.10 | 98.13±0.09 | **99.20±0.06** |
| halfd-thres | 98.27±0.09 | 98.97±0.07 | 97.96±0.10 | 98.32±0.09 | **99.22±0.06** |

| 99%-FPR | $L_2$ | $L_1$ | SymmKL | $\chi^2$ | GCL |
|---|---|---|---|---|---|
| trevi-l2 | 11.36±1.65 | 3.44±0.75 | 8.02±1.04 | 8.02±1.08 | **2.42±0.58** |
| trevi-thres | 7.14±1.31 | 3.24±0.69 | 7.93±1.11 | 5.06±0.97 | **2.23±0.48** |
| notre-l2 | 19.69±1.93 | 6.09±0.72 | 14.81±1.66 | 9.40±1.04 | **4.16±0.57** |
| notre-thres | 11.9±1.19 | 5.17±0.58 | 13.11±1.39 | 8.24±1.12 | **3.72±0.56** |
| halfd-l2 | 44.55±9.42 | 34.01±2.10 | 43.51±1.07 | 40.53±1.12 | **26.06±2.25** |
| halfd-thres | 40.58±1.63 | 32.30±2.28 | 42.51±1.22 | 39.28±1.49 | **26.36±2.50** |

Table 1: The average precision (above) and the false positive rate at 99% recall (below) of different distance measures on the Photo Tourism datasets, without jitter effects.