

# Patch matching with polynomial exponential families and projective divergences

Frank Nielsen<sup>1</sup> and Richard Nock<sup>2</sup>

<sup>1</sup> École Polytechnique and Sony Computer Science Laboratories Inc  
Frank.Nielsen@acm.org

<sup>2</sup> Data61 & the Australian National University, Australia

**Abstract.** Given a query patch image, patch matching consists in finding similar patches in a target image. In pattern recognition, patch matching is a fundamental task that is time consuming, specially when zoom factors and symmetries are handled. The matching results heavily depend on the underlying notion of distances, or similarities, between patches. We present a method that consists in modeling patches by flexible statistical parametric distributions called polynomial exponential families (PEFs). PEFs model universally arbitrary smooth distributions, and yield a compact patch representation of complexity independent of the patch sizes. Since PEFs have computationally intractable normalization terms, we estimate PEFs with score matching, and consider a projective distance: the symmetrized  $\gamma$ -divergence. We demonstrate experimentally the performance of our patch matching system.

## 1 Introduction and prior work

Given a query patch image  $I_s$  of dimension  $(w_s, h_s)$ , patch matching asks to find “similar” patches in a target image  $I_t$  of dimension  $(w_t, h_t)$ . Patch matching find countless applications [1–3] in image processing. A basic dissimilarity measure of patches  $I_s$  and sub-image patch  $I_t(x_0)$  of dimension  $(w_s, h_s)$  anchored at location  $x_0$  is the Sum of Square Differences (SSDs) of the pixel intensities:  $D(I_s, I_t(x_0)) = \sum_{x \in [1, w_s] \times [1, h_s]} (I_s(x) - I_t(x + x_0))^2$ , that can be interpreted as the squared Euclidean distance on the vectorized patch intensities. Thus finding similar patches amount to find *close(st) neighbors* in  $\mathbb{R}^{w_s \times h_s}$ . This basic SSD distance may be further extended to color or multi-channel images (like hyperspectral images) either by taking the average or the maximum of the SSDs for all channels. A naïve brute-force baseline patch matching algorithm computes a matching score for each potential pixel position  $x_0 \in [1, w_t] \times [1, h_t]$  at the target image in time  $O(w_t h_t w_s h_s)$ . This is prohibitively too expensive in practice. When dealing with pure translations, the Fourier phase correlation method [4, 5] can be used to speed up the alignment of images within subpixel precision in  $O(wh \log(wh))$  time using the Fast Fourier Transform (FFT) with  $w = \max(w_s, w_t)$  and  $h = \max(h_s, h_t)$ . To factorize and speed up the distance calculation between the source patch and the target patch when scanning the target image, a general Patch Matching (PM) method [1, 2] has been designed

that computes a Nearest Neighbor Field (NNF). However, those methods are too time consuming when dealing with large patches, and are not robust to smooth patch deformations, symmetry detections (like reflections [6, 7]) of patches, and zooming factors (requiring guessing and rescaling the source patch accordingly).

We propose a fast statistical framework to match patches: Our algorithm models (potentially large) patches by statistical parametric distributions estimated from patch pixels, and define the distance between patches by a corresponding statistical distance between those compact patch representations. In order to handle a flexible faithful modeling of any arbitrary smooth probability density, we consider Polynomial Exponential Families [8, 9] (PEFs) that have intractable normalizing constants. Since we cannot computationally normalize those PEF distributions, we bypass this problem by considering statistical projective distances [10] that ensures that  $D(\lambda q, \lambda' q') = D(q, q')$  for any  $\lambda, \lambda' > 0$ , where  $q$  and  $q'$  are the unnormalized distributions of patches.

The paper is organized as follows: Section 2 presents our patch statistical representation, the chosen distance function, and the fast batched patch parameter estimation procedure using integral images. Section 3 reports on our experiments. Section 4 concludes this work by hinting at further perspectives.

## 2 Patch matching with polynomial exponential families

### 2.1 Polynomial Exponential Families: Definition and estimation

We consider the *univariate* parametric probability distributions with the following probability density:

$$p(x; \theta) = \exp(\langle \theta, t(x) \rangle - F(\theta)),$$

where  $t(x) = (x^1, \dots, x^D)$  denotes the sufficient statistics [11, 10], and  $\theta$  the natural parameters.  $\langle x, y \rangle = x^\top y \in \mathbb{R}$  denotes the Euclidean inner product. Let  $\Theta \subseteq \mathbb{R}^D$  be the natural parameter space so that  $\int \exp(\langle \theta, t(x) \rangle) dx < \infty$ .

The function  $F(\theta) = \log \int \exp(\langle \theta, t(x) \rangle) dx$  is called the log-normalizer or partition function in statistical physics [11, 10]. The order of this exponential family is the dimension of the natural parameter space,  $D$ . Since  $t_i(x) = x^i$  is the monomial of degree  $i$  for  $i \in [D] = \{1, \dots, D\}$ , this family of distribution is called a *polynomial exponential family* (PEF). PEFs are *universal* density estimators that allow one to model arbitrarily finely any smooth multimodal density [8]. This can be easily seen by considering the log-density that is a polynomial function, and polynomial functions are well-known to model any smooth function.

For an exponential family, the maximum likelihood estimator  $\hat{\theta}$  from a set of independently and identically distributed (iid) scalar observations  $x_1, \dots, x_n$  (the patch pixel intensities with  $n = w_s h_s$ ) satisfies the following identity equation:  $\nabla F(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n t(x_i)$ .

Since  $F(\theta)$  is *not* known in closed form for PEFs, one cannot compute its gradient  $\nabla F(\theta)$ , and we need a different method to estimate  $\hat{\theta}$ . Let  $q(x; \theta) =$

$\exp(\langle \theta, t(x) \rangle)$  be the unnormalized model, a positive probability measure (and  $p(x; \theta) = \frac{q(x; \theta)}{e^{F(\theta)}} \propto q(x; \theta)$ ). We use the *Score Matching* Estimator [12, 13] (SME<sup>3</sup>):

$$\hat{\theta} = - \left( \sum_{i=1}^n D(x_i)^\top D(x_i) \right)^{-1} \left( \sum_{i=1}^n \Delta t(x_i) \right), \quad (1)$$

where  $D(x) = \nabla t(x) = (t'_1(x), \dots, t'_D(x))$  is the vector of derivatives of  $t(x)$  (term by term), and  $\Delta t(x)$  is the Laplacian operator, (computed from the second derivatives, term by term). We have  $t'_i(x) = ix^{i-1}$  when  $i \geq 1$  (and 0 otherwise) and  $t''_i(x) = i(i-1)x^{i-2}$  when  $i \geq 2$  (and 0 otherwise). Notice that SME is not efficient when  $p(x; \theta)$  is not Gaussian [10] (but MLE is efficient). See also [8] for alternative estimation recursion moment method of PEFs. Here, we consider  $\mathcal{X} = \mathbb{R}^+$  the support of all PEFs (although intensity values are clamped to 255 for fully saturated pixels).

Thus a patch of size  $(w_s, h_s)$  is represented *compactly* by a natural parameter of a PEF of order  $D \ll w_s \times h_s$ , and is *independent* of the patch resolutions.

## 2.2 Statistical projective divergences

In order to measure the (dis)similarity between two patches described by their natural parameters  $\theta_s$  and  $\theta_t$ , we need a proper *statistical distance* [14–16] like the relative entropy also called the Kullback-Leibler (KL) divergence:  $\text{KL}(p(x; \theta_s), p(x; \theta_t)) = \int_{x \in \mathbb{R}^+} p(x; \theta_s) \log \frac{p(x; \theta_s)}{p(x; \theta_t)} dx$ . It is well-known that the KL divergence amounts to a Bregman divergence on swapped natural parameters when the distributions come from the same exponential family [17]. However, we point out that we do not have the log-normalizer  $F(\theta)$  in closed form for PEFs. Hence, we consider a *projective divergence* that ensures that  $D(\lambda q, \lambda' q') = D(q, q')$  for any  $\lambda, \lambda' > 0$ . For PEFs, we need to consider a statistical projective distance, and we choose the  $\gamma$ -divergence [18–20] (for  $\gamma > 0$ ) between two distributions  $p$  and  $q$ :

$$D_\gamma(p, q) = \frac{1}{\gamma(1+\gamma)} \log I_\gamma(p, p) - \frac{1}{\gamma} \log I_\gamma(p, q) + \frac{1}{1+\gamma} \log I_\gamma(q, q), \quad (2)$$

where

$$I_\gamma(p, q) = \int_{x \in \mathcal{X}} p(x) q(x)^\gamma dx. \quad (3)$$

When  $\gamma \rightarrow 0$ ,  $D_\gamma(p, q) \rightarrow \text{KL}(p, q)$ . For our patch matching application, we consider the *symmetrized  $\gamma$ -divergence*:  $S_\gamma(p, q) = \frac{1}{2}(D_\gamma(p, q) + D_\gamma(q, p))$ . Although  $D_\gamma(p, q)$  can be applied to unnormalized densities, its value depend on the log-normalizer  $F$ . Indeed, the term  $I_\gamma(p : q)$  admits a closed-form solution provided that  $\gamma\theta_p + \theta_q \in \Theta$ :

$$I_\gamma(\theta_p, \theta_q) = \exp(F(\theta_p + \gamma\theta_q) - F(\theta_p) - \gamma F(\theta_q)). \quad (4)$$

<sup>3</sup> [http://user2015.math.aau.dk/presentations/invited\\_steffen\\_lauritzen.pdf](http://user2015.math.aau.dk/presentations/invited_steffen_lauritzen.pdf)

*Proof.* We have  $I_\gamma(\theta_p, \theta_q) = \int \exp(\langle t(x), \theta_p + \gamma\theta_q \rangle - F(\theta_p) - \gamma F(\theta_q)) dx$ . Expanding the right-hand side, we get  $\exp(F(\theta_p + \gamma\theta_q) - F(\theta_p) - \gamma F(\theta_q)) \int \exp(\langle t(x), \theta_p + \gamma\theta_q \rangle - F(\theta_p + \gamma\theta_q)) dx$ . By definition, when  $\gamma\theta_p + \theta_q \in \Theta$ , we have  $\int \exp(\langle t(x), \theta_p + \gamma\theta_q \rangle - F(\theta_p + \gamma\theta_q)) dx = 1$ , and the result follows. The  $\gamma\theta_p + \theta_q \in \Theta$  condition is always satisfied when the natural parameter space [15, 16, 14] is a *cone* (since  $\gamma > 0$ ), like the multivariate Gaussians distributions, the multinomial distributions, and the Wishart distributions, just to name a few.

One can check that by taking a Taylor expansion on the  $\gamma$ -divergence expressed using the closed-form expression of Eq. 4 for exponential families with conic natural space, we obtain the Bregman divergence [17] when  $\gamma \rightarrow 0$ .

To fix ideas, we shall consider  $\gamma = 0.1$  in the remainder. Since the support is univariate, we may approximate the  $\gamma$ -divergence by discretizing the integral of Eq. 3 with a *Riemann sum* (discretization). Another method that works also for arbitrary multivariate distributions, is to use *stochastic integration* by sampling  $x_1, \dots, x_m$  following distribution  $p$  (importance sampling). Then we have:

$$I_\gamma(p, q) = \int_{x \in \mathcal{X}} p(x)q(x)^\gamma dx \simeq \frac{1}{m} \sum_{i=1}^m q(x_i)^\gamma.$$

In practice, we set  $m = 100000$  for importance sampling.

Notice that some statistical divergences are only *one-sided* projective divergence. For example, Hyvärinen divergence [21]:

$$D_H(p, q) = \frac{1}{2} \int p(x) \|\nabla_x \log p(x) - \nabla_x \log q(x)\|^2 dx.$$

### 2.3 Fast PEF estimations using Summed Area Tables

Recall that to compute the score matching estimator of Eq. 1 we need to compute both sums for *all*  $w_t \times h_t$  locii of the patches. In order to estimate the PEF parameters in constant time (for a prescribed order  $D$ ) instead of time proportional to the source patch size, we use *Summed Area Tables* [22] (SATs also called integral images, interestingly also used to detect mirror symmetry of patches in [6]). For every pixel of position  $(x, y)$ , the value of the summed area table  $F(x, y)$  is  $\sum_{x' \leq x, y' \leq y} f(x', y')$ , where  $f$  is the function that we want to sum up. In our setting, we need two SATs (cumulative sum arrays) per channel for  $D^\top D$  and  $\Delta t$  evaluated for the intensity (or red, green, blue values when considering color images). The value of the table at position  $(x, y)$  can be computed using previously computed values when filling the SAT,

$$F(x, y) = f(x, y) + F(x - 1, y) + F(x, y - 1) - F(x - 1, y - 1)$$

(when  $x - 1$  or  $y - 1$  is zero, the value is just zero). Once the table is constructed, for the score matching method, when we want to compute the sums in the equation for a patch (a rectangle with bottom-left corner of position  $(x_0, y_0)$  and top-right corner of position  $(x_1, y_1)$ ), and we compute it in  $O(1)$  as:

$$\sum_{x_0 \leq x \leq x_1, y_0 \leq y \leq y_1} f(x, y) = F(x_1, y_1) + F(x_0, y_0) - F(x_0, y_1) - F(x_1, y_0).$$

### 3 Experiments

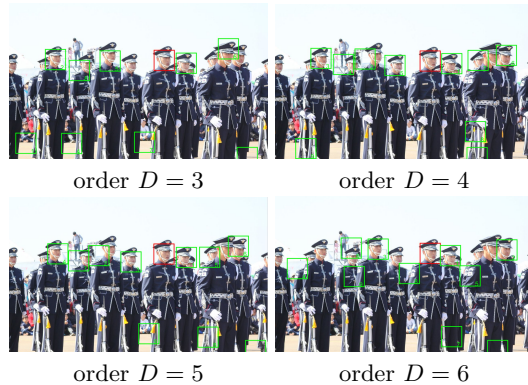


**Fig. 1.** Comparison of PEFPM with the baseline SSD patch matching (patch size  $150 \times 150$  and image size  $960 \times 640$ ): Observe that due to its flexibility the statistical modeling got more face hits (4 faces) than the pixel-aligned SSD method (one face). The upper row is on pixel intensities, the middle row on sum of RGB channel dissimilarities, and the last row on the maximum of RGB channel dissimilarity.

All our experiments were performed in Java<sup>TM</sup> on a HP Elitebook 840 G1 (i7-4600U CPU 2.1 GHz with 8 GB RAM). First, let us compare our PEFPM method (with  $\gamma = 0.1$ ) with the baseline SSD method: Figure 1 displays the results obtained when considering intensity values, sum of distances (sum of SSDs or sum of  $S_\gamma$  divergences), or max of distances for the dissimilarity between patches. Observe that our statistical method successfully detected 4 visually similar patches (human faces) while the aligned pixel-based distance detected only one face.

Next, we study the impact of the PEF order  $D$  on the visual patch search in Figure 2. We notice that results depend on the chosen order, and that the most visually similar patches are found for  $D = 4$  and  $D = 5$ . This raises the problem of *model selection* for future research.

Table 1 reports the computational times broken down into (i) the SAT construction stage, (ii) the PEF estimations, and (iii) the  $S_\gamma$  approximation according to the order of the PEF. Clearly, the higher the order the more costly. The PEF estimation stage scales linearly with the order with slope 1, while the SAT construction and  $S_\gamma$  search have slopes  $< 1$ .

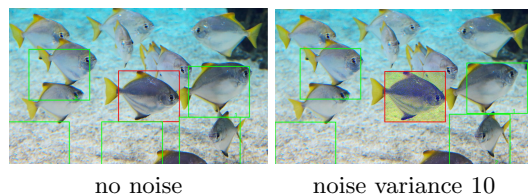


**Fig. 2.** Impact of the polynomial degree of PEFs, the order  $D$  of the exponential family (image size  $1280 \times 853$ , patch size  $100 \times 100$ ). Here, order  $D = 4$  and  $D = 5$  find the most visually similar patches.

	degree of PEF				
	2	3	4	5	6
SAT construction	5.99	6.37	8.04	9.56	11.86
PEF estimation	5.24	7.44	8.96	12.84	14.82
$S_\gamma$ search	7.02	8.86	9.70	10.30	10.52

**Table 1.** Computational time in seconds for PEFs.

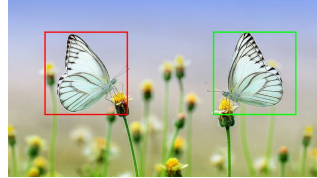
We then tested the *stability* of our statistical PEFPM method by adding some gaussian noise (Figure 3). Corrupting the pixel channels with a Gaussian noise change the estimated distributions, but the distance evaluations with the gamma divergence provably attenuate the distortions [18], and we can check that we obtained the same matching patches.



**Fig. 3.** Effect of Gaussian noise on PEFs patch matching (patch size  $300 \times 250$ ).

Finally, one big advantage of our statistical modeling other pixel-based patch distance is to allow to consider different zoom values and symmetries. Indeed, geometric symmetries do not change (much) the estimated distribution in a patch. For example, see Figure 4 that illustrates this property: PEFPM nicely detected the two butterflies even if one looks like the reflected copy of the other.

Here, the patch sizes are large (about 1/3 of the image dimension) and SSD-based method will be very costly and inefficient.



**Fig. 4.** Patch matching with symmetry (reflection) detected by PEFPM of order 6 (patch size  $250 \times 250$ , image dimension  $(960, 640)$ ).

## 4 Concluding remarks

We proposed a novel statistical flexible modeling of image patches for fast pattern recognition. Our approach is particularly well-suited for handling large patches and accounting for patch symmetries [7] or other deformations in target images. This work offers many avenues for future research: (i) *model selection* of the polynomial exponential family according to the query patch, (2) *foreground/background* detection in patches (say, using Grabcut [23]) and matching only the foreground statistical distributions to improve the accuracy of patch matching, (3) *multivariate PEFs* to bypass sum/max of univariate PEF distances, etc.

## Acknowledgments

We gratefully thank Quei-An Chen (École Polytechnique, France) for implementing our patch matching system and performing various experiments.

## References

1. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)* **28**(3) (2009) 24
2. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized patch-match correspondence algorithm. In: *European Conference on Computer Vision (ECCV)*. Springer (2010) 29–43
3. Li, Y., Dong, W., Shi, G., Xie, X.: Learning parametric distributions for image super-resolution: Where patch matching meets sparse coding. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. (Dec 2015) 450–458
4. Kuglin, C.: The phase correlation image alignment method. In: *Proc. Int. Conf. on Cybernetics and Society, 1975*. (1975) 163–165

5. Foroosh, H., Zerubia, J.B., Berthod, M.: Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing* **11**(3) (2002) 188–200
6. Patraucean, V., Gioi, R., Ovsjanikov, M.: Detection of mirror-symmetric image patches. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. (2013) 211–216
7. Wang, Z., Tang, Z., Zhang, X.: Reflection symmetry detection using locally affine invariant edge correspondence. *IEEE Transactions on Image Processing* **24**(4) (April 2015) 1297–1301
8. Cobb, L., Koppstein, P., Chen, N.H.: Estimation and moment recursion relations for multimodal distributions of the exponential family. *Journal of the American Statistical Association* **78**(381) (1983) 124–130
9. Rohde, D., Corcoran, J.: MCMC methods for univariate exponential family models with intractable normalization constants. In: *Statistical Signal Processing (SSP), 2014 IEEE Workshop on*. (June 2014) 356–359
10. Amari, S.i.: *Information Geometry and Its Applications*. Applied Mathematical Sciences. Springer Japan (2016)
11. Brown, L.D.: Fundamentals of statistical exponential families with applications in statistical decision theory. *Lecture Notes-monograph series* **9** (1986) i–279
12. Hyvärinen, A.: Estimation of non-normalized statistical models by score matching. In: *Journal of Machine Learning Research*. (2005) 695–709
13. Forbes, P.G., Lauritzen, S.: Linear estimating equations for exponential families with application to gaussian linear concentration models. *Linear Algebra and its Applications* **473** (2015) 261 – 283 Special issue on Statistics.
14. Nielsen, F., Nock, R.: A closed-form expression for the Sharma-Mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical* **45**(3) (2011) 032003
15. Nielsen, F.: Closed-form information-theoretic divergences for statistical mixtures. In: *Pattern Recognition (ICPR), 21st International Conference on, IEEE* (2012) 1723–1726
16. Nielsen, F., Nock, R.: On the chi square and higher-order chi distances for approximating  $f$ -divergences. *IEEE Signal Processing Letters* **1**(21) (2014) 10–13
17. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with Bregman divergences. *The Journal of Machine Learning Research* **6** (2005) 1705–1749
18. Fujisawa, H., Eguchi, S.: Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis* **99**(9) (2008) 2053–2081
19. Notsu, A., Komori, O., Eguchi, S.: Spontaneous clustering via minimum gamma-divergence. *Neural computation* **26**(2) (2014) 421–448
20. Chen, T.L., Hsieh, D.N., Hung, H., Tu, I.P., Wu, P.S., Wu, Y.M., Chang, W.H., Huang, S.Y., et al.:  $\gamma$ -sup: A clustering algorithm for cryo-electron microscopy images of asymmetric particles. *The Annals of Applied Statistics* **8**(1) (2014) 259–285
21. Ehm, W.: Unbiased risk estimation and scoring rules. *Comptes Rendus Mathématiques* **349**(11) (2011) 699–702
22. Crow, F.C.: Summed-area tables for texture mapping. *ACM SIGGRAPH computer graphics* **18**(3) (1984) 207–212
23. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)* **23**(3) (2004) 309–314