

High-Order Statistics for Point Prediction in Natural Images

Wilson S. Geisler and Jeffrey S. Perry

Center for Perceptual Systems

1 University Station

University of Texas at Austin

Austin TX, 78712

geisler@psy.utexas.edu

jsp@mail.utexas.edu

Abstract: Results are presented for a simple conditional-moments method that directly measures high-order statistics of natural images. In four estimation tasks significant increases in performance are obtained in comparison to traditional methods.

OCIS codes: (100.0100) Image processing; (110.0110) Imaging systems; (330.0330) Vision, color, and visual optics

1. Introduction

Characterizing the structure of natural images is critical for understanding visual encoding and decoding in biological vision systems and for applications in image processing and computer vision. The most studied statistical regularities in natural images involve first- and second-order (pair-wise) statistics [1,2]. Further, it is typical to impose some invariance constraints in order to reduce dimensionality or otherwise simplify measurement of the statistical regularities. However, restricting measurements to pair-wise statistics and making invariance assumptions may miss important statistical structure [3]. Here we apply the simple classical approach of measuring moments along single dimensions, conditional on the values along other dimensions. We apply this approach to measuring high-order local image statistics relevant to the four tasks shown in Fig. 1. Although the conditional moments approach is only practical for distributions of modest dimension, it has some unique advantages. First, univariate conditional distributions are frequently unimodal and simple in shape, and thus the first few moments capture much of the shape information. Second, estimating conditional moments only requires keeping a single running sum for each moment, making it practical to use essentially arbitrarily large numbers of training signals and hence to measure high order statistics with higher precision. Third, the conditional first moment (conditional mean) is the Bayes optimal estimator when the cost function is the mean squared error (i.e., the MMSE estimator). Fourth, conditional moments can be measured recursively in a hierarchical fashion, allowing the approach to be extended to higher numbers of dimensions than would otherwise be practical.

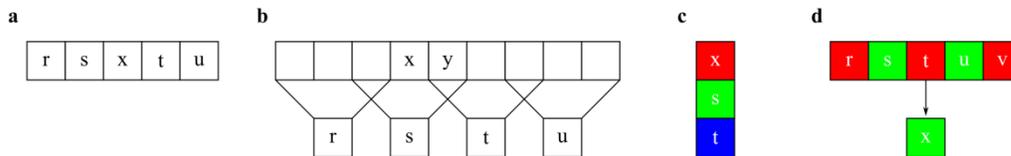


Fig. 1. Four estimation tasks. **a.** Estimation of a missing image point from neighboring points. **b.** Estimation of a high resolution image from a low resolution image (super resolution). **c.** Estimation of a missing color channel without taking into account spatial information. **d.** Estimation of a missing color filter array pixel (demosaicing). Letters *x* and *y* represent values to be estimated; other letters are observed values.

2. Methods

The image set consisted of 1049 images of outdoor scenes collected with a calibrated Nikon D700 camera. The images contained no man-made objects and the exposure for each image was set so as to minimize pixel response saturation (clipping). Each of the 4284 x 2844 raw images was converted to linear 8-bit gray scale or 24-bit color. For some measured statistics the images were first converted to LMS cone space. The images were randomly shuffled and then divided into two groups: 700 training images that were used to generate the tables, and 349 test images. All pixels in the training images were used to estimate conditional moments, and thus the number of training samples was on the order of 10^{10} .

Under most circumstances, an arbitrary joint probability density function $p(\mathbf{x}_n)$ can be specified by measuring its single dimensional conditional moments, which can be estimated simply by summing the observed values of x_i^k for each value of the vector \mathbf{x}_{i-1} , and then dividing by the number of times the vector \mathbf{x}_{i-1} was observed. Thus, $E(x_i^k | \mathbf{x}_{i-1}) \cong \sum_{x_i \in \Omega(\mathbf{x}_{i-1})} x_i^k / N(\mathbf{x}_{i-1})$, where $\mathbf{x}_{i-1} = (x_1, \dots, x_{i-1})$, $\Omega(\mathbf{x}_{i-1})$ is the set of observed values of x_i for the

IMC1.pdf

given value of the vector \mathbf{x}_{i-1} , and $N(\mathbf{x}_{i-1})$ is the number of those observed values. We define the optimal estimate function for x_i to be $f(\mathbf{x}_{i-1}) = E(x_i | \mathbf{x}_{i-1})$, which gives the Bayes optimal estimate of x_i when the goal is to minimize the mean squared error of the estimated value from the true value [4]. It is possible to obtain estimates based on even higher-order statistics by recursively measuring conditional first moments given the lower-order estimates and their relative reliability obtained from the conditional second moments. For example, let \hat{x}_1 be the optimal estimate of x_i given one set of conditional variables, let \hat{x}_2 be the optimal estimate of x_i given a different set of conditional variables, and let \hat{r}_{12} be their estimated relative reliability, then one can obtain an estimate given all the conditional variables by measuring $g(\hat{x}_1, \hat{x}_2, \hat{r}_{12}) = E(x_i | \hat{x}_1, \hat{x}_2, \hat{r}_{12})$.

3. Results

3.1 Estimating missing image points. If only the closest two pixels s and t were used to estimate x , and if the joint probability distribution $p(x, s, t)$ were Gaussian, then the best estimate of x would be the linear-regression estimate obtained from the pair-wise correlations. Fig. 2a plots the difference between the optimal and linear-regression estimates. The plot not only reveals the differences from the Gaussian prediction, it shows that higher-order structure would be lost if simple invariance assumptions were made. Specifically, the plot shows that the optimal estimate of the missing pixel relies crucially upon the specific values of s and t and not upon their relative values. For example, if one were to normalize each signal prior to measuring the statistics (subtract the mean of the three points from each point or convert to a contrast signal), then much of the structure in Fig. 2a would be missed.

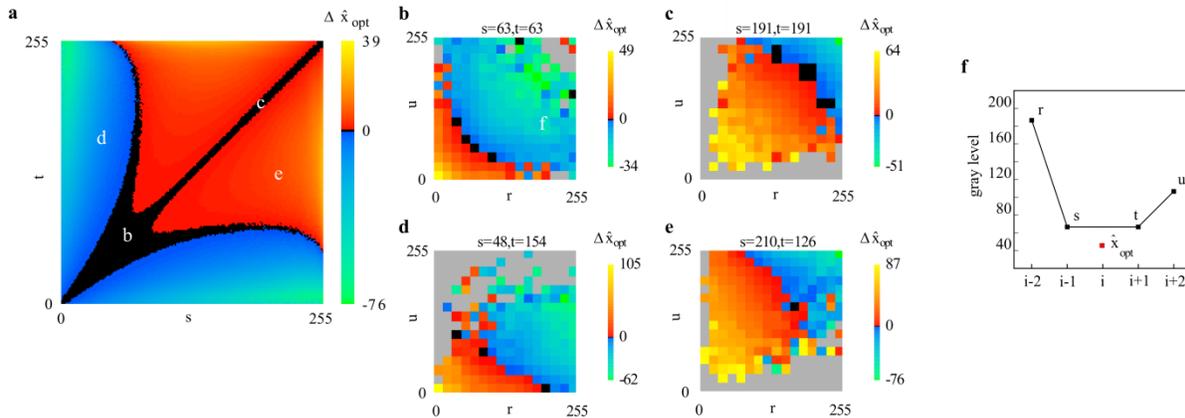


Fig. 2. Estimation of a missing or occluded point. **a.** Third-order statistics: optimal estimation of x given (s, t) . The color map shows the difference between the optimal and linear-regression estimate. **b-e.** Fifth-order statistics: optimal estimation of x given (r, s, t, u) , for the specific values of (s, t) indicated by the white letters in **a**; $b=(63, 63)$, $c=(191, 191)$, $d=(48, 154)$, $e=(210, 126)$. Each plot gives the difference between the optimal estimate and average of s and t as a function of the further flanking values r and u . Gray indicates no data. **f.** Example of a specific point prediction indicated by the white letter in **b**. The black symbols show the observed values (r, s, t, u) and the red symbol shows the optimal prediction. In this figure the histogram plots are unsmoothed.

Now consider estimating a missing image point from four of its neighbors. These fifth-order statistics can be visualized as 65536 (256×256) two-dimensional plots. Four instances of these 65536 plots are shown in figures 2b-e. Each plot corresponds to a letter marked in Fig. 2a. The values in these plots were quantized into 16×16 bins. The plots are systematic and complex, and they are also consistent with expectations. For example, Fig. 2f shows an instance of a data point in Fig. 2b. When s and t are equal, third order statistics predict that x equals s and t , but when r and u are included, the optimal prediction of x is less than s and t . This is consistent with what one would expect in a smoothly varying image profile. For the test images, the average PSNR improvement per image over bilinear estimation was 3.2 dB (3 dB corresponds to a 50% reduction in mean squared error). We also note that even better performance is obtained by optimally combining vertical and horizontal estimates. The unoptimized, non-parallelized version of the interpolation code can process images at a rate of 137 machine-cycles/pixel.

3.2 Estimating a higher resolution image from a lower resolution image (super resolution). Fig. 3a shows an example of the fourth-order statistics relevant for super resolution (i.e., estimating x given r, s, t in Fig. 1b). Plotted is the difference between the optimal estimate and the bilinear estimate when the value of s is 63. Note that this is one of 255 such plots. As above, the statistics are complex but systematic. The average PSNR improvement per image over bilinear, spline (MATLAB interp2) and cubic (MATLAB interp2) upsampling was 3.5 dB, 1.7 dB and

IMC1.pdf

2.3 dB, respectively. The unoptimized, non-parallelized upsampling application processes at a rate of 176 machine-cycles/pixel.

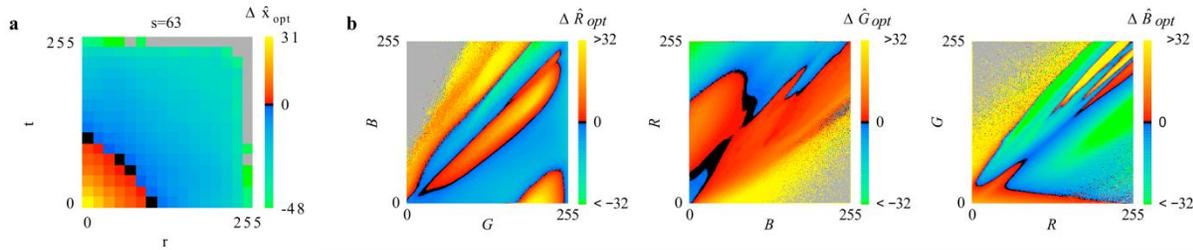


Fig. 3 Upsampling and color channel estimation. **a.** Estimation of higher resolution image. Optimal estimation of x given (r, t) when $s=63$. The color indicates the difference between the optimal estimate of x and s . **b.** Estimation of missing R , G or B camera responses. Plots show the difference between the optimal estimate and the estimate based upon linear regression. In all plots, gray indicates no data. In this figure the histogram plots are unsmoothed.

3.3 Estimating missing color channels. Fig. 3b plots the third-order statistics relevant for estimating a missing R , G , or B color channel. The plots for estimating L , M , and S cone responses are topologically similar to these plots and have similar complexity and regularity.

The estimation error was calculated by removing a color channel from the image and then calculating the mean squared error between the estimated color and the original color at each pixel in the perceptually uniform CIE $L^*a^*b^*$ color-space. The average increase in PSNR of the optimal estimate over the linear regression estimate was $(R, G, B) = (1.2, 0.59, 0.93)$ dB. For LMS cone space the average increase in PSNR was $(L, M, S) = (1.3, 0.9, 2.3)$ dB. Note also that in almost all images, at least one of the optimally restored R , G , or B channel (or L , M , or S response) images had natural looking color balances, whereas most of the images restored using linear regression contained unnatural tints in each of the restored channel images.

3.4 Estimating of a missing color filter array pixel (demosaiicing). For comparison purposes we chose a state-of-the-art algorithm, AHD. The optimal estimate was calculated by using the closest two green and three red pixels along a single dimension neighboring the green pixel that we wished to estimate, as shown in Fig. 1d. Applying AHD to the test images resulted in a PSNR of 35.8 dB. Applying the optimal estimates to the same pixels results in a PSNR of 36.6 dB. Also, AHD performed 3 times slower. This is a preliminary finding for only the green channel.

4. Conclusion

The direct conditional-moments method has many advantages. Firstly, because the learning portion of the method relies upon summing the values of the training pixels, it scales linearly with the number of pixels in the training set, and thus the sets may be arbitrarily large. Secondly, the estimation process involves only table lookups and is therefore very efficient. Our preliminary interpolation and upsampling codes require only one or two hundred machine cycles per pixel, and our parallelized codes can process over 10 twelve megabyte images per second. Additionally, the method can be applied separably in both dimensions, making it practical for hardware that can operate upon only a few image scanlines at a time. The statistics also have implications for biological vision. For example, the statistics in Fig. 2 make specific predictions for human performance in estimating missing pixel values [5], the statistics in Fig. 3a make testable predictions for differences between foveal and peripheral vision, and the statistics for estimating missing color filter pixels are relevant for the problem of interpolating cone responses [6,7].

5. Acknowledgements

Supported by NIH grant EY11747.

6. References

- [1] Simoncelli EP, Olshausen BA. (2001) Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24:1193–216.
- [2] Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. of Psych.*, 59, 167-192.
- [3] Lee AB, Pedersen KS & Mumford D (2003) The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision* 54(1/2/3), 83–103.
- [4] Bishop, C.M. (2006) *Pattern recognition and machine learning*. Springer, New York NY.
- [5] Kersten, D. (1987) Predictability and redundancy of natural images. *J. Opt. Soc. Am. A*, 4, 2395-2400.
- [6] Hofer H, Singer B & Williams DR (2005) Different sensations from cones with the same photopigment. *Journal of Vision*, 5(5), 444-454.
- [7] Brainard, DH, Williams, DR & Hofer H (2008) Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *Journal of Vision*, 8(5):15, 1–23.