Vision Research 50 (2010) 2213–2222

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Lower bounds on the redundancy of natural images

Reshad Hosseini, Fabian Sinz, Matthias Bethge *

Max Planck Institute for Biological Cybernetics, Spemannstraße 41, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 23 October 2009
Received in revised form 28 July 2010

Keywords:
Redundancy
Natural image statistics
Multi-information rate
Information theory

ABSTRACT

The light intensities of natural images exhibit a high degree of redundancy. Knowing the exact amount of their statistical dependencies is important for biological vision as well as compression and coding applications but estimating the total amount of redundancy, the multi-information, is intrinsically hard. The common approach is to estimate the multi-information for patches of increasing sizes and divide by the number of pixels. Here, we show that the limiting value of this sequence—the multi-information rate—can be better estimated by using another limiting process based on measuring the *mutual information* between a pixel and a causal neighborhood of increasing size around it. Although in principle this method has been known for decades, its superiority for estimating the multi-information rate of natural images has not been fully exploited yet. Either method provides a lower bound on the multi-information rate, but the *mutual information* based sequence converges much faster to the multi-information rate than the conventional method does. Using this fact, we provide improved estimates of the multi-information rate of natural images and a better understanding of its underlying spatial structure.

1. Introduction

Natural images contain an abundance of structure and regularities which can be quantified as statistical dependencies or redundancy between image pixels. Coding and compression algorithms for photographic images exploit these dependencies for achieving a good performance. Besides technical applications, the statistical regularities in natural images also play an important role for our understanding of sensory coding in the mammalian brain. In a wide range of studies it has been shown that many response properties of neurons in the early visual system such as color opponency, bandpass filtering, contrast gain control and orientation selectivity can be interpreted as mechanisms for removing these redundancies in natural images (Atick & Redlich, 1992; Barlow, 1959; Buchsbaum & Gottschalk, 1983; Karklin & Lewicki, 2008; Linsker, 1990; Olshausen & Field, 1996; Schwartz & Simoncelli, 2001; Simoncelli & Olshausen, 2001; Sinz & Bethge, 2009; Srinivasan, Laughlin, & Dubs, 1982). Quantitative comparisons have shown that these response properties are not all equally effective in removing statistical dependencies. Mechanisms removing second-order correlations in natural images such as color opponency and bandpass filtering yield a large reduction of redundancy. Less pronounced but still substantial is the effect of contrast gain control (Lyu & Simoncelli, 2009; Sinz & Bethge, 2009). For orientation selectivity, however, the potential for redundancy reduction turns out to be much smaller (Bethge, 2006). Since the emergence of orientation selectivity is the most prominent difference in the response properties of V1 neurons compared to the retina it can serve as an important witness on whether neural response properties in cortex can still be interpreted convincingly in terms of redundancy reduction (Eichhorn, Sinz, & Bethge, 2009).

An important unknown that is critical to judging this case is the true total amount of redundancy in natural images. A principled way of quantifying redundancy is to measure the *multi-information* of a distribution (Perez, 1977). The multi-information of a multivariate random variable is the difference between the sum of its marginal entropies and its joint entropy

$$I[X_1 : \ldots : X_n] = \sum_{i=1}^{n} H[X_i] - H[X_1, \ldots, X_n].$$

It equals zero if and only if the individual components are statistically independent and is positive otherwise. It measures the information gain caused by statistical dependencies between the single variables. Unlike differential entropy, the multi-information is invariant against arbitrary component-wise transformations both for linear mappings, such as scaling, and nonlinear mappings, such as taking the logarithm.

The conventional approach for estimating the redundancy per pixel—the *multi-information rate*—is to estimate the multi-information for patches of increasing sizes and divide by the number of pixels (Bethge, 2006; Chandler & Field, 2007; Eichhorn et al., 2009; Lee, Wachtler, & Sejnowski, 2002; Lewicki & Olshausen, 1999; Lewicki & Sejnowski, 2000; Lyu & Simoncelli, 2009; Sinz & Bethge, 2009; Wachtler, Lee, & Sejnowski, 2001). In this way we

* Corresponding author.
E-mail address: mbethge@tuebingen.mpg.de (M. Bethge).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.07.025

obtain a monotonically increasing sequence converging to the multi-information rate

$$I_\infty = \lim_{n\to\infty} \frac{1}{n} I[X_1 : \ldots : X_n].$$

There is an important trade-off between two different kinds of errors that affect the outcome of this limiting process: On the one hand, the earlier we stop the sequence of increasing patch sizes, the more we ignore long-range dependencies between image pixels and, hence, underestimate the redundancy of natural images. On the other hand, the larger the patch sizes get, the more difficult it becomes to estimate the multi-information reliably due to the increase in dimensionality. Multi-information estimation strongly resembles the problem of estimating the joint density and similarly suffers from the curse of dimensionality: The number of states that need to be estimated grows exponentially with the number of dimensions. This means that more and more regularization is needed to avoid overfitting in high dimensions. As a consequence, with increasing dimensionality it becomes increasingly unlikely to capture all the structure of the density.

The trade-off between ignoring long range correlations for small n and the increasing difficulty to estimate $I[X_1:\ldots:X_n]$ for large n suggests that the estimation of the multi-information rate can be improved substantially if one manages to construct sequences other than $\{\frac{1}{n}I[X_1 : \ldots : X_n]\}_{n=1}^{\infty}$ which converge faster to the same limiting value I_∞.

In this paper, we show that it is possible to construct such a sequence. The basic idea can be illustrated in the case of one-dimensional stationary stochastic processes. From information theory it is known that the conditional entropy converges to the entropy rate of such processes[1] (Cover & Thomas, 2006; Shannon, 1948)

$$\lim_{n\to\infty} \frac{1}{n} H[X_1, \ldots, X_n] = \lim_{n\to\infty} H[X_n|X_{n-1}, \ldots, X_1].$$

Multiplying this equation by (-1) and adding the marginal entropy of the stationary process $H[X_1] = \frac{1}{n}\sum_{k=1}^{n} H[X_k]$ at both sides, yields an analogous relationship for the multi-information rate

$$\begin{aligned} I_\infty &= \lim_{n\to\infty} \frac{1}{n} I[X_1 : \ldots : X_n] = \lim_{n\to\infty} I[X_n : X_{n-1}, \ldots, X_1] \\ &= \lim_{n\to\infty} H[X_n] - H[X_n|X_{n-1}, \ldots, X_1]. \end{aligned} \tag{1}$$

Note that the sequence on the left hand side of Eq. (1) reflects the *multi*-information[2] between all the variables $X_1, \ldots, X_n$ while the sequence on the right hand side reflects the *mutual* information between X_n and $(X_1, \ldots, X_{n-1})$. The mutual information is the special case of the multi-information which measures the statistical dependencies between two random variables only, while it is possible that the dimensionality of the two random variables is different. For example, in our case X_n is a univariate random variable and $(X_1, \ldots, X_{n-1})$ is $(n-1)$-dimensional. The chain rule for the multi-information (Cover & Thomas, 2006)

$$I[X_1 : \ldots : X_n] = \sum_{k=2}^{n} I[X_k : X_{k-1}, \ldots, X_1],$$

shows that the multi-information can be decomposed into a sum of mutual information terms. This suggests that the mutual information based sequence $\{I_n^{inc}\}_{n=1}^{\infty}$ with $I_n^{inc} := I[X_n : X_{n-1}, \ldots, X_1]$ quantifies the asymptotic *increment* in the multi-information while the conventionally used multi-information based sequence $\{I_n^{cum}\}_{n=1}^{\infty}$ with $I_n^{cum} := \frac{1}{n} I[X_n : \ldots : X_1]$ constitutes a *cumulative* approach which averages over these increments.

[1] For continuous random variables it is necessary to additionally assume that the limit exists.

[2] More precisely the multi-information divided by n.

Inspired by an early study in the fifties (Schreiber, 1956), an incremental approach for estimating I_∞ has already been used before in Petrov and Zhaoping (2003) but did not reveal its full potential. Our work elucidates a couple of points that have not been addressed in those papers: First, we revise the mathematical justification for using the incremental approach in case of two-dimensional random fields rather than one-dimensional processes as it is necessary for modeling images. Second, we show that the mutual information based method yields significantly better estimates of I_∞ than the conventional method does while Petrov and Zhaoping (2003) did not provide any comparisons with previous methods. Third, we show how particularly reliable multi-information estimators can be constructed for the incremental approach such that one obtains conservative lower bounds to the multi-information rate. This allows us, fourth, to systematically investigate how the two approaches perform on natural images for different number of dimensions n also far beyond the case of $n = 7$ pixels that was studied in Petrov and Zhaoping (2003). Our best lower bound on the multi-information rate for the van Hateren data set exceeds their estimate by more than 20% and slightly outperforms the bound obtained with the L_p-spherical model (Sinz & Bethge, 2009). It is obtained when using a causal neighborhood of only 25 pixels.

The remaining part of the paper is structured as follows: In Section 2, we introduce the *multi-information* based and the *mutual information* based method for estimating the multi-information rate. In particular, we present a proof for the convergence of the two methods to the same limiting value I_∞ for two-dimensional stationary stochastic processes. In Section 3, we perform experiments on artificial images in order to demonstrate the validity of the method, and apply it to natural images afterwards. Our results show that the incremental method based on conditional distributions performs significantly better and indicates that the multi-information rate of natural images contains a substantial contribution from higher-order moments. We further corroborate this finding by a second set of experiments where we first pre-whiten the images before we fit the local image statistics. In this way, we not only confirm our previous estimates for the multi-information rate but we can also show that the predominant statistical dependencies captured by current models of natural images are of very limited spatial extent. In particular, the increase in the multi-information rate observed for the cumulative method for increasing patch size does not reflect a meaningful contribution of long range correlations but rather an artifact caused by the pixels at the boundary. Finally, in Section 5, we discuss the significance of our results and compare them to existing work.

2. Methods

In order to describe the statistical regularities of natural images, they are often modeled as two-dimensional stationary random fields. For the present study, stationarity is crucial as it is provides the critical link between the *cumulative* and the *incremental* method for computing the multi-information rate. Stationarity means that the random field is invariant under translations with respect to the x- and y-coordinates of the image intensities. In the following, we will first depict the mathematical underpinnings for using the incremental approach in case of two-dimensional stationary random fields. After that we will show that the incremental method is generally superior to the cumulative method, and then we will describe how to construct reliable multi-information and mutual information estimators for the cumulative and the incremental method, respectively. In particular, we will construct conservative estimators such that also the empirical quantities become reliable lower bounds to the multi-information rate.