Dipole statistics of discrete finite images: two visually motivated representation theorems

Charles Chubb and John I. Yellott

Department of Cognitive Sciences, University of California at Irvine, Irvine, California 92697-5100

Received August 27, 2001; revised manuscript received October 12, 2001; accepted October 12, 2001

A discrete finite image I is a function assigning colors to a finite, rectangular array of discrete pixels. A dipole is a triple, $((d_R, d_C), \alpha, \beta)$, where d_R and d_C are vertical and horizontal, integer-valued displacements and α and β are colors. For any such dipole, $D_I((d_R, d_C), \alpha, \beta)$ gives the number of pixel pairs $((r_1, c_1), (r_2, c_2))$ of I such that $I[r_1, c_1] = \alpha$, $I[r_2, c_2] = \beta$ and $(r_2, c_2) - (r_1, c_1) = (d_R, d_C)$. The function D_I is called the dipole histogram of I. The information directly encoded by the image I is purely *locational*, in the sense that I assigns colors to locations in space. By contrast, the information directly encoded by D_I is purely *relational*, in the sense that D_I registers only the frequencies with which pairs of intensities stand in various spatial relations. Previously we showed that any discrete, finite image I is uniquely determined by D_I [Vision Res. **40**, 485 (2000)]. The visual relevance of dipole histogram representations is questionable, however, for at least two reasons: (1) Even when an image viewed by the eye nominally contains only a small number of discrete color values, photon noise and the random nature of photon absorption in photoreceptors imply that the effective neural image will contain a far greater (and unknown) range of values, and (2) D_I is generally of much greater cardinality than I. First we introduce "soft" dipole representations, which forgo the perfect registration of intensity implicit in the definition of D_I, and show that such soft representations uniquely determine the images to which they correspond; then we demonstrate that there exists a relatively small dipole representation of any image. Specifically, we prove that for any discrete finite image I with $N > 1$ pixels, there always exists a restriction Q of D_I (with the domain of Q dependent on I) of cardinality at most $N - 1$ sufficient to uniquely determine I, provided that one also knows N; thus there always exists a purely relational representation of I whose order of complexity is no greater than that of I itself.

OCIS codes: 100.2960, 100.5010, 330.7310, 330.5000.

1. INTRODUCTION

A discrete finite image $I[r, c]$ is a function that assigns real numbers (representing colors) to a finite, rectangular array of ordered pairs (r, c) of integers (representing spatial locations). Thus I directly codes purely *locational* information. For many purposes, however, locational information is of little direct visual utility. For example, locational information is irrelevant for identifying a face or recognizing an object or assessing a property of some surface texture. More important for such purposes are the spatial relationships between intensities in the visual field. Indeed, it is commonly found that neurons (in inferotemporal cortex) subserving such purposes have large receptive fields and are typically quite tolerant to large changes of location of stimuli within their receptive fields.[1–4]

Here we explore some basic properties of relational image representations. This is a class of image-coding schemes that may prove useful for pattern recognition and other visual judgments that depend primarily on sensitivity to internal relationships between components of a pattern. Moreover, the representations in this class lend themselves gracefully to size-invariant assessments as well. (Face-selective neurons in monkey STS are remarkably size invariant.[5,6] Indeed, in their study of 33 such face-selective neurons, Rolls and Baylis[6] found that the median size change tolerated with a response of greater than half the maximal response was 12 times.)

A basic example of a relational representation is the dipole histogram D_I of an image I. A dipole is a triple, $((d_R, d_C), \alpha, \beta)$, where d_R and d_C are horizontal and vertical, integer-valued displacements and α and β are real numbers (color values). In any such dipole, α is called the initial value, and β is called the terminal value. For any such dipole, $D_I((d_R, d_C), \alpha, \beta)$ gives the number of pixel pairs $((r_1, c_1), (r_2, c_2))$ of I such that $I[r_1, c_1] = \alpha$, $I[r_2, c_2] = \beta$, and $(r_2, c_2) - (r_1, c_1) = (d_R, d_C)$. (Thus D_I has the obligate symmetry, $D_I((d_R, d_C), \alpha, \beta) = D_I((-d_R, -d_C), \beta, \alpha)$.) Note that the information directly encoded by D_I is exclusively relational.

The dipole histogram of an image and the related notion of the probabilistic, second-order statistics of a stochastic image have played an important role in past discussions of texture segregation.[7–17] However, dipoles have not played an important role in recent computational models of texture processing.[18–22]

It is easy to see how D_I can be constructed from I. Although it may not be quite so obvious, I can also be uniquely constructed from D_I.[23] Thus any image I can be represented either in terms of purely locational information or in terms of the purely relational information embodied in D_I.

The purely relational nature of the information directly coded in the dipole histogram makes it an intriguing alternative image representation. However, there are obvious reasons to suppose that the visual system does not in fact construct dipole histograms of its input. First, the dipole histogram is typically a much larger object than the corresponding image—that is, the cardinality of D_I tends to be much greater than that of the original input image I (especially in the case in which most of I's pixel values are distinct). Second, although we can imagine artificial images comprising only a few different discrete colors, the images encountered in daily life typically comprise continuously varying colors, almost none of which are likely to be identical between different pixels within a given image. Moreover, even when an image viewed by the eye nominally contains only a small number of discrete color values, photon noise and the random nature of photon absorption in photoreceptors imply that the effective neural image will contain a far greater (and unknown) range of values. To compute the dipole histogram of such an image requires a level of precision in intensity resolution that we have no reason to think is achieved by the visual system.

In this paper, after basic definitions in Section 2, we shall begin in Section 3 by generalizing the notion of the dipole histogram of an image I to that of a "soft" dipole representation of I. Soft dipole representations relax the requirement (implicit in the definition of the dipole histogram) that intensities within an image be precisely registered, allowing instead a "softened" or "graded" registration of intensity. In this respect, at least, soft dipole representations may claim greater biological plausibility than dipole histograms. Moreover, as we shall show, precise registration of image intensity can be sacrificed with no loss of information: A soft dipole representation of a given input image uniquely determines that image.

After introducing soft dipole representations, and showing that they uniquely determine the images to which they correspond, we return in Section 4 to (nonsoft) dipole histograms. Our purpose in Section 4 is to show that relational representations need not be exorbitantly large compared with the images to which they correspond. Specifically, we show that for any image I (of any dimension) comprising $N > 1$ pixels, there always exists a restriction (term defined in Subsection 4.A) Q_I of D_I to a subset H_I of at most $N - 1$ dipoles with the following property: For any image J comprising N pixels, if $D_J(\chi) = Q_I(\chi)$ for all dipoles $\chi \in H_I$, then $J = I$. In other words, the restriction Q_I uniquely determines I (provided that one knows the number of pixels in I).

In summary, our aim is twofold: First, we will show that the notion of the dipole histogram can be generalized to a family of representations that do not require precise registration of intensity; second, we will show that relational representations (specifically, dipole representations) need not be huge in comparison with the images they represent. We hope that these two results will stimulate further research in this area.

2. ONE- AND TWO-DIMENSIONAL IMAGES AND THEIR DIPOLE HISTOGRAMS

A. One-Dimensional Images

Let $R = \{0, 1, \ldots, N - 1\}$, for some integer N. Then, a one-dimensional image is a function $I: R \rightarrow \Re$. Elements of R are called pixels, and for any pixel r, $I[r]$ denotes the value assigned by I to r.

A one-dimensional dipole is a triple, (d, α, β), with d a nonnegative integer-valued displacement and α and β real numbers. We say that a dipole (d, α, β) bridges a pair (r_1, r_2) of pixels in I if $r_2 - r_1 = d$, $I[r_1] = \alpha$, and $I[r_2] = \beta$. The dipole histogram D_I assigns to each dipole $\chi = (d, \alpha, \beta)$ the number of distinct pairs in I bridged by χ. Thus, if $D_I(6, 1, -4.3) = 16$, then there are 16 pixels r of I such that $I[r] = 1$ and $I[r + 6] = -4.3$. We write Support(D_I) for the set of all dipoles χ such that $D_I(\chi) > 0$.

B. Two-Dimensional Images

Let $R = \{0, 1, \ldots, N - 1\}$ and $C = \{0, 1, \ldots, M - 1\}$ for integers N and M. A two-dimensional image is a function $J: R \times C \rightarrow \Re$, where $R \times C$ denotes the Cartesian product of R and C. The elements of $R \times C$ are called pixels, and $J[r, c]$ denotes the value assigned by J to a given pixel (r, c).

A two-dimensional dipole is a triple, (d, α, β), with displacement $d = (d_R, d_C)$ comprising a nonnegative, integer-valued vertical displacement d_R and a (possibly negative) integer-valued, horizontal displacement d_C. (We need to allow either negative row or column displacements in order to capture all varieties of two-dimensional displacement, each in one direction; as a matter of convention, we let column displacements vary in sign.)

As in the one-dimensional case, α and β are real numbers. We say that a dipole (d, α, β) bridges a pair $[(r_1, c_1), (r_2, c_2)]$ of pixels in J if $d = (r_2, c_2) - (r_1, c_1)$, $J[r_1, c_1] = \alpha$ and $J[r_2, c_2] = \beta$. As in the one-dimensional case, the dipole histogram D_J assigns to each dipole $\chi = (d, \alpha, \beta)$ the number of distinct pairs of pixels in J bridged by χ, and we continue to write Support(D_J) for the set of all dipoles χ such that $D_J(\chi) > 0$.

3. SOFT DIPOLE REPRESENTATIONS

One reason to question the biological relevance of dipole histogram representations is the following: Even when an image viewed by the eye nominally contains only a small number of discrete color values, photon noise and the random nature of photon absorption in photoreceptors imply that the effective neural image will contain a far greater (and unknown) range of values. In this section we explore the possibility of relaxing the assumption implicit in the definition of the dipole histogram that image intensities are precisely registered by the visual system.

To take an example, let I be an image, d a displacement, and α and β intensities. Then $D_I(d, \alpha, \beta)$ gives the total number of pixels r for which $I[r] = \alpha$ and $I[r + d] = \beta$. By contrast, let neighborhood(α) and neighborhood(β) be the open intervals $(\alpha - 1/2, \alpha + 1/2)$ and $(\beta - 1/2, \beta + 1/2)$, and consider the function $S_I(d, \alpha, \beta)$ giving the total number of pixels r such that $I[r] \in$ neighborhood(α) while $I[r + d] \in$ neighborhood(β). S_I is an example of a soft dipole representation. As is evident in its definition, S_I embodies information similar to that embodied by D_I; however, S_I forgoes the precise registration of image intensities implicit in the definition of D_I. Nonetheless, as we show here, I is uniquely determined by S_I.

A. Soft Dipole Representations of One-Dimensional Images

It will be convenient to obtain some basic results about one-dimensional images and then extend them to higher-dimensional cases. Accordingly, let $G:\Re \to \Re$ be a probability density function with mean 0. Then, for any one-dimensional image I with N pixels, any $\alpha, \beta \in \Re$, and $d = 0, 1, \ldots, N-1$, define

$$S_I(d, \alpha, \beta) = \sum_{r=0}^{N-1-d} G(I[r] - \alpha)G(I[r+d] - \beta). \tag{1}$$

[Note that the sum in Eq. (1) ranges over all pixels r for which $I[r]$ and $I[r+d]$ are both defined.]

Suppose, for example, that

$$G(v) = \begin{cases} 1 & \text{if } -\dfrac{1}{2} < v < \dfrac{1}{2}. \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In this case, as discussed above, S_I behaves as an intensity-tolerant version of D_I. The crucial difference is as follows: $D_I(d, \alpha, \beta)$ counts precisely those occurrences of pixels r for which $I[r] = \alpha$ and $I[r+d] = \beta$; by contrast, S_I does not require strict equality between $I[r]$ and α or between $I[r+d]$ and β: for any given α and β, $S_I(d, \alpha, \beta)$ gives the total number of pixels r in $\mathrm{Dom}(I)$ such that $I[r]$ is within distance 1/2 of α and $I[r+d]$ is within distance 1/2 of β.

Another natural choice for G is a Gaussian density with mean 0:

$$G(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp -(v^2/2\sigma^2). \tag{3}$$

With G given by Eq. (3), it is not as easy to give a precise verbal characterization of $S_I(d, \alpha, \beta)$ [as with G defined by Eq. (2)]. It is clear, however, that as long as σ is relatively small in comparison with the range of pixel values occurring in I, then $S_I(d, \alpha, \beta)$ continues to reflect the degree to which I tends to assign values near α to pixels r while assigning values near β to pixels $r + d$.

The choices of G given in Eqs. (2) and (3) yield definitions of S_I similar in spirit to D_I. However, there are many choices of G for which this is not true. For example, one might choose a bimodal (zero-mean) density G with $G(0) = 0$. Such a choice of G would yield an S_I largely insensitive to the information reflected by D_I. Specifically, $S_I(d, \alpha, \beta)$ would be utterly uninfluenced by occurrences of pixels r for which $I[r] = \alpha$ and $I[r+d] = \beta$—precisely the events that $D_I(d, \alpha, \beta)$ counts.

However, regardless of the value that G assigns to 0, if G is a density function with mean 0, then I is uniquely determined by S_I. Note, for example, that

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & S_I(d, \alpha, \beta)\alpha \mathrm{d}\alpha \mathrm{d}\beta \\ &= \sum_{r=0}^{N-1-d} \int_{-\infty}^{\infty} G(I[r] - \alpha)\alpha \mathrm{d}\alpha \int_{-\infty}^{\infty} G(I[r+d] - \beta)\mathrm{d}\beta \\ &= \sum_{r=0}^{N-1-d} I[r],\end{aligned} \tag{4}$$

implying that one way of retrieving I from S_I is to take

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & [S_I(N-1-d, \alpha, \beta) - S_I(N-d, \alpha, \beta)]\alpha \\ & \times \mathrm{d}\alpha \mathrm{d}\beta = \sum_{r=0}^{d} I[r] - \sum_{r=0}^{d-1} I[r] = I[d].\end{aligned} \tag{5}$$

The reader will note, however, that although S_I uniquely determines I, S_I is uncountable in cardinality, being continuous in each of α and β. In general, finite approximations of S_I will correspond only to approximations of I.

B. Soft Dipole Representations of Two-Dimensional Images

Let I be a two-dimensional image with N rows and M columns. Then [by analogy to Eq. (1)] define

$$\begin{aligned}S_I((d_r, d_c), \alpha, \beta) = \sum_{r=0}^{N-1-d_r} \sum_{c=0}^{M-1-d_c} & G(I[r, c] - \alpha) \\ & \times G(I[r+d_r, c+d_c] - \beta).\end{aligned} \tag{6}$$

Note that in Eq. (6), d_r is constrained to be in the range $\{0, 1, \ldots, N-1\}$, while d_c is allowed to take all values in $\{1-M, 2-M, \ldots, -1, 0, 1, \ldots, M-1\}$.

Let J be the one-dimensional image obtained by concatenating the successive rows of I. Specifically,

$$\begin{aligned} J[rM + c] &= I[r, c], \\ r &= 0, 1, \ldots, N-1, \\ c &= 0, 1, \ldots, M-1. \end{aligned} \tag{7}$$

Note, then, that any dipole $((d_r, d_c), \alpha, \beta)$ of I corresponds uniquely to the dipole $(d_r M + d_c, \alpha, \beta)$ of J. Moreover, for any dipole (d, α, β) of J, there exist at most two dipoles of I that correspond to (d, α, β). If they both exist, these two dipoles are $((d_r, d_c), \alpha, \beta)$ and $((d_r + 1, d_c - M), \alpha, \beta)$, where d_r is the greatest integer such that $d_r M \leqslant d$ and $d_c = d - d_r M$. If either $d_c = 0$, or $d_r = N - 1$, then $((d_r + 1, d_c - M), \alpha, \beta)$ will be undefined, in which case only the dipole $((d_r, d_c), \alpha, \beta)$ of I will correspond to dipole (d, α, β) of J. In this case, reflection on Eqs. (6) and (1) shows that

$$S_J(d, \alpha, \beta) = S_I((d_r M, d_c), \alpha, \beta), \tag{8}$$

whereas otherwise,

$$\begin{aligned} S_J(d, \alpha, \beta) = {} & S_I((d_r M, d_c), \alpha, \beta) \\ & + S_I(((d_r + 1)M, d_c - M), \alpha, \beta). \end{aligned} \tag{9}$$

Thus S_I uniquely determines S_J. Equation (5) thus implies that J is uniquely determined by S_I. In addition, we can infer the number M of columns in I by surveying Support(S_I): Specifically, M is equal to 1 plus the maximum column displacement in any dipole in Support(S_I). This final item of information extracted from S_I makes possible the construction of I. We need only assign the first M pixel values of J to the first row of I, the second M pixel values of J to the second row of I, etc.

This shows that any two-dimensional image I is uniquely determined by the corresponding soft dipole representation S_I. Moreover, it should be clear that this argument generalizes straightforwardly to images of any dimension.

C. Possible Roles of Soft Dipole Representations in Visual Processing

Equation (5) shows that soft dipole representations preserve complete information about the input image I even though they sacrifice precise registration of image intensity. However, for any given displacement d, $S_I(d, \alpha, \beta)$ varies continuously over all $\alpha, \beta \in \Re$. Moreover, the method used in Eq. (5) to reconstruct I from S_I makes use of the entire continuum of values $S_I(d, \alpha, \beta)$, $d = 0, 1, \ldots, N-1$, and $\alpha, \beta \in \Re$. This might be taken to suggest that infinitely many neurons would be required to fully encode the information in I (even though I itself comprises only $N+1$ pixels).

Several comments are in order. First (and most important), it is unlikely that visual mechanisms using soft dipole representations would require those representations to preserve complete information about the input. Recognition of a face, for example, might well depend on the extraction of only a portion of the relational information inherent in the input image [e.g., only the values $S_I(d, \alpha, \beta)$ corresponding to a restricted set of crucially informative dipoles, (d, α, β)].

Second, for natural choices of G [e.g., G given by Eq. (3)], reflection on Eq. (1) suggests that for any given displacement d, the function $S_I(d, \alpha, \beta)$ will vary slowly in each of the arguments α and β. For example, if G is a Gaussian function, we can expect $S_I(d, \alpha, \beta)$ to be low-pass filtered in each of α and β and hence band limited. In this case, however, it seems likely that nearly all the information in I may be preserved by a relatively low-dimensional representation of S_I.

This issue remains to be explored in detail. We shall, however, address a closely related question in the next section.

4. SMALL DIPOLE REPRESENTATIONS

Recall that any function is a set of ordered pairs. This is true in particular of D_I although each individual element of D_I is rather complicated: An element of D_I is an ordered pair $(\chi, D_I(\chi))$, where $\chi = ((d_R, d_C), \alpha, \beta)$ is a dipole and $D_I(\chi)$ is the number of occurrences of χ in I. We call any subset $Q \subseteq D_I$ a dipole representation of I if the following condition holds: For any image J, if J has the same number of pixels as I, and $Q \subseteq D_J$, then $J = I$. In this case, we shall sometimes (equivalently) say that Q uniquely determines I.

The cardinality of any finite set is simply the number of elements in that set. Thus, for a dipole representation Q of some finite image I, the cardinality of Q (denoted $|Q|$) is the number of ordered pairs $(\chi, D_I(\chi))$ in Q. Note, however, that a given dipole χ can occur in at most one ordered pair of Q; thus $|Q|$ gives the number of dipoles on which Q is defined.

For any image I, the dipole histogram D_I is itself a dipole representation of I.[23] However, D_I is typically greater in cardinality than the original image I. For example, consider an image I with N rows by M columns of pixels that assigns a different color to each pixel. In this special case, every dipole occurring in I differs from every other dipole in either its initial or its terminal value. If we limit redundancy in D_I by allowing only displacements $d = (d_R, d_C)$ such that $d_R \geqslant 0$ and $d_C \geqslant 0$ whenever $d_R = 0$, then each pair of (possibly identical) pixel locations in I corresponds to a single type of dipole of I. Thus the set of dipoles over which D_I is nonzero is of cardinality $N_{\text{pixels}}(N_{\text{pixels}}+1)/2$ for $N_{\text{pixels}} = NM$; and for each dipole $((d_R, d_C), \alpha, \beta)$ in this set, $D_I((d_R, d_C), \alpha, \beta) = 1$.

Our purpose in this section is to show that for any discrete, finite image I comprising $N > 1$ pixels there always exists a dipole representation of I of cardinality at most $N-1$ (provided that one also knows the number of pixels in I). The proof we offer is constructive; that is, we show how to produce such a representation for any given image I. The representation we construct is not necessarily the smallest that can be obtained (though in some cases it is). Moreover, we make no claim that the particular representation we derive is biologically plausible or useful. Our proof merely establishes an upper bound on the size of a minimal representation.

A. Proper Subsets of D_I That Uniquely Determine I

Chubb and Yellott[23] showed that any one-dimensional image I is uniquely determined by D_I (and extended this result to images of arbitrary dimensionality). We include a simple proof of this fact (in the one-dimensional case) in Appendix A. However, it often turns out that D_I is highly redundant: We can usually find proper subsets of D_I that uniquely determine I.

Let H be a subset of Support(D_I), and define $Q: H \to \Re$ by setting $Q(\chi) = D_I(\chi)$ for any dipole $\chi \in H$. The subset Q of D_I is called the restriction of D_I to H. For a given image I, there often exist proper subsets of D_I that suffice to determine all the pixel values of I. In other words, it often turns out that D_I is redundant.

To take an example, let I be a one-dimensional image all of whose pixels are assigned distinct values. It is easy to see that I will be uniquely determined by the restriction Q of D_I to the set H of dipoles (d, α, β) such that $d = 1$. There will exist in H a unique dipole $(1, \alpha_1, \beta_1)$ such that α_1 appears as the terminal value in none of the other dipoles of H. It is easy to see that $I[1] = \alpha_1$, and that $I[2] = \beta_1$. Moreover, there will be a unique dipole $(1, \alpha_2, \beta_2) \in H$ such that $\alpha_2 = \beta_1$. Clearly, $I[3] = \beta_2$, and so forth, until all of I's pixel values are determined. Of course, we must also be given the information that I has N pixels; otherwise, we cannot rule out the possibility that some of I's dipoles with displacement 1 are missing from H.

To take another simple example, suppose all of I's pixels take the same value, γ. In this case, Support(D_I) comprises only the N dipoles, $(0, \gamma, \gamma)$, $(1, \gamma, \gamma)$,..., $(N - 1, \gamma, \gamma)$. However, D_I is nonetheless highly redundant. Indeed, I is uniquely determined by the restriction Q of D_I to the singleton $\{(0, \gamma, \gamma)\}$. The knowledge that $Q(0, \gamma, \gamma) = N$ in conjunction with the information that I has N pixels tells us immediately that all of I's pixels take value γ. In this case, then, the minimal dipole representation of I is smaller in cardinality than I itself.

In both of the examples we have considered, we exploited special properties of the given image I to extract a restriction of D_I sufficient to determine I. The main point of this section is to show that special properties are not required. A relatively small restriction of D_I can always be derived that suffices to determine I. The proof that we offer is constructive and addresses directly the multidimensional case.

B. Multidimensional Images

Let $N_M, N_{M-1},\ldots, N_0$ be natural numbers; and for $i = M, M-1,\ldots, 0$, let $X_i = \{0, 1,\ldots, N_i - 1\}$. Then call $X = X_M \times X_{M-1} \times \ldots \times X_0$ the [$(M+1)$-dimensional] pixel lattice with dimensions $N_M, N_{M-1},\ldots, N_0$ (where $X_M \times X_{M-1} \times \ldots \times X_0$ denotes the Cartesian product of sets $X_M, X_{M-1},\ldots, X_0$). Then any function $I{:}X \to \mathfrak{R}$ is called an $(M+1)$-dimensional image. Any vector $x \in X$ is called a pixel of I, and we write $I[x]$ for the value assigned to x by I.

An $(M+1)$-dimensional dipole is a triple, (d, α, β), with initial and terminal values $\alpha, \beta \in \mathfrak{R}$ and displacement $d = (d_M, d_{M-1},\ldots, d_0)$ comprising integer coordinate values. We say that a dipole (d, α, β) bridges a pair (x, y) of pixels in I if $d = y - x$, $I[x] = \alpha$ and $I[y] = \beta$. As in the one-dimensional case, the dipole histogram D_I assigns to each dipole $\chi = (d, \alpha, \beta)$ the number of distinct pairs of pixels in I bridged by χ; we continue to write Support(D_I) for the set of all dipoles χ such that $D_I(\chi) > 0$.

C. There Exists a Relatively Small Dipole Representation of Any Image

The proof of the main result of this Subsection (Proposition 4.C.2) makes use of two functions, λ and its inverse π, which we now define. Note that the pixel lattice X comprises

$$N = N_M N_{M-1} \ldots N_0 \tag{10}$$

pixels. We assign integers $0, 1,\ldots, N-1$ to the pixels in X by setting

$$\lambda[x] = x_M\left(\prod_{i=0}^{M-1} N_i\right) + x_{M-1}\left(\prod_{i=0}^{M-2} N_i\right) + \ldots + x_1 N_0 + x_0 \tag{11}$$

for any pixel $x = (x_M, x_{M-1},\ldots, x_0) \in X$.

For any pixels $x, y \in X$, one quickly sees that $\lambda[y] \geq \lambda[x]$ iff y lexicographically dominates x. It is also clear that λ is 1–1 from X onto $\{0, 1,\ldots, N-1\}$, implying that the inverse function λ^{-1} is well defined. It will be convenient to write π for λ^{-1}; thus, for any $n \in \{0, 1,\ldots, N-1\}$, $\pi(n)$ is the nth pixel of X under lexicographic ordering.

A few observations will facilitate the proof of Proposition 4.C.2.

1. *Lemmas*

1. For any $x, y \in X$,

(a) If $x + y \in X$, then $\lambda[x+y] = \lambda[x] + \lambda[y]$. (12)

(b) If $y - x \in X$, then $\lambda[y-x] = \lambda[y] - \lambda[x]$. (13)

To derive (a), note that the vector $x + y$ will be an element of X only if $x_i + y_i \leq N_i - 1$, for $i = 0, 1,\ldots, M$. In this case, the fact that Eq. (11) is linear in x implies that $\lambda[x+y] = \lambda[x] + \lambda[y]$. The argument for (b) is similar.

2. For any $n \in \{0, 1,\ldots, N-1\}$,

$$\pi[n] - \pi[0] = \pi[n] = \pi[N-1] - \pi[N-1-n]. \tag{14}$$

The left-hand equality follows directly from the fact that $\pi[0] = (0, 0,\ldots, 0)$; to derive the right-hand equality, note that $y = \pi[N-1] = (N_M - 1, N_{M-1} - 1,\ldots, N_0)$ and $x = \pi[N-1-n]$ satisfy the conditions of Eq. (13) above. Thus

$$\begin{aligned}
\pi[N-1] - \pi[N-1-n] &= \pi[\lambda[\pi[N-1] - \pi[N-1-n]]] \\
&= \pi[\lambda[\pi[N-1]] - \lambda[\pi[N-1-n]]] \\
&= \pi[N-1-[N-1-n]] \\
&= \pi[n].
\end{aligned} \tag{15}$$

3. For any $x, y \in X$,

(a) If $\lambda[x] + \lambda[y] > N - 1$, then $x + y \notin X$. (16)

(b) If $\lambda[y] - \lambda[x] < 0$, then $y - x \notin X$. (17)

To see that (a) holds, derive the contrapositive: Suppose that $x + y \in X$. Then [by Eq. (12)] $\lambda[x] + \lambda[y] = \lambda[x + y] \leq N - 1$. Similarly, to see that (b) holds, note that if $(y - x) \in X$, then [by Eq. (13)] $\lambda[y] - \lambda[x] = \lambda[y - x] \geq 0$.

We are now equipped to prove the main result of this section.

2. *Proposition*

For some positive integer M, let I be an image on the pixel lattice X with dimensions $N_M, N_{M-1},\ldots, M_0$. Then there exists a dipole representation of I of cardinality at most $N - 1$, where $N = N_M N_{M-1} \ldots N_0$.

Proof. For simplicity, suppose that N is even; that is, suppose that $N = 2K$ for some integer K. Then let

$$H_I = \left\{ \begin{array}{ll} (\pi[N-1], I[\pi[0]], I[\pi[N-1]]), & \\ (\pi[N-2], I[\pi[0]], I[\pi[N-2]]), & (\pi[N-2], I[\pi[1]], I[\pi[N-1]]), \\ (\pi[N-3], I[\pi[0]], I[\pi[N-3]]), & (\pi[N-3], I[\pi[2]], I[\pi[N-1]]), \\ \vdots & \vdots \\ (\pi[K], I[\pi[0]], I[\pi[K]]), & (\pi[K], I[\pi[K-1]], I[\pi[N-1]]) \end{array} \right\}. \quad (18)$$

(In Eq. (18) we take the liberty of writing $\pi[k]$ sometimes to denote the pixel x for which $\lambda[x] = k$ and at other times to denote the displacement d whose offsets are equal to the coordinate values of x.) Note first that $|H_I|$ is *at most* $N - 1$. $|H_I|$ can be less than $N - 1$ if dipoles $(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]])$ and $(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]])$ happen to be identical, for one or more j (i.e., if $I[\pi[0]] = I[\pi[j]]$ and $I[\pi[N-j-1]] = I[\pi[N-1]]$). We want to show that the restriction Q_I of D_I to H_I is a dipole representation of I. Accordingly, let J be an image with N pixels, and suppose that $Q_I \subset D_J$.

Immediately we note that the dimensionality of J is determined by Q_I: Given that the total number of pixels in J is N, the fact that the displacement of the first dipole in H_I is $(N_M - 1, N_{M-1} - 1, \ldots, N_0 - 1)$ implies that J has dimensions $N_M, N_{M-1}, \ldots, N_0$.

J's dimensionality implies in turn that J has only one dipole with displacement $\pi[N-1] = (N_M - 1, N_{M-1} - 1, \ldots, N_0 - 1)$: the dipole that bridges pixels $(0, 0, \ldots, 0)$ and $(N_M - 1, N_{M-1} - 1, \ldots, N_0 - 1)$. Thus, from the fact that $D_J(\pi[N-1], I[\pi[0]], I[\pi[N-1]]) = Q_I(\pi[N-1], I[\pi[0]], I[\pi[N-1]])$, we infer that $J[\pi[0]] = I[\pi[0]]$ and $J[\pi[N-1]] = I[\pi[N-1]]$.

We proceed by induction (on j). Suppose that for some integer $0 < j < K$, $J[\pi[i]] = I[\pi[i]]$ for $i = 0, 1, \ldots, j-1$ and also for $i = N-1, N-2, \ldots, N-j$. We shall show that in this case, it is also true that $J[\pi[j]] = I[\pi[j]]$ and $J[\pi[N-1-j]] = I[\pi[N-1-j]]$.

The argument begins with a demonstration that for almost all pairs x, y of pixels in X separated by displacement $\pi[N-1-j]$, both of $J[x]$ and $J[y]$ are determined by the induction hypothesis. Specifically, we shall show that there are only two pairs of pixels x_1, y_1 and x_2, y_2 separated by displacement $\pi[N-1-j]$ for which this is not true: The first such pair has $x_1 = \pi[0]$ and $y_1 = \pi[N-1-j]$ (y_1 is undetermined); the second such pair has $x_2 = \pi[j]$ and $y_2 = \pi[N-1]$ (x_2 is undetermined). Note that for each of these two pixel pairs Eq. (14) above implies that the displacement from x to y is $\pi[N-1-j]$.

A necessary (but insufficient) condition for pixels x and y to be bridged by a dipole with displacement $\pi[N-1-j]$ is that

$$\text{(i)}\ \lambda[x] \leqslant j \quad \text{and} \quad \text{(ii)}\ \lambda[y] \geqslant N-1-j. \quad (19)$$

Argument: If (i) is false (i.e., if $\lambda[x] > j$), then $\lambda[x] + N - 1 - j > N - 1$, in which case Eq. (16) above implies that the vector $x + \pi[N-1-j]$ is not in the pixel lattice X. Similarly, if (ii) is false, then $\lambda[y] < N - 1 - j$, in which case Eq. (17) implies that $y - \pi[N-1-j]$ is not in X.

Consider then some $x \in X$ for which $0 < \lambda[x] < j$, and let $y = x + \pi[N-1-j]$. It is possible that y may not be a pixel in X. However, by Eq. (12), if $y \in X$, then $\lambda[y] = \lambda[x] + N - 1 - j > N - 1 - j$. In this case, the induction hypothesis implies that $J[x] = I[x]$ and $J[y] = I[y]$. Thus, as asserted at the outset, there exist only two pairs of pixels $x_1, y_1 \in X$ and $x_2, y_2 \in X$ separated by displacement $\pi[N-1-j]$ for which the induction hypothesis does not directly suffice to determine that $J[x] = I[x]$ and $J[y] = I[y]$. The first such pair has $x_1 = \pi[0]$ and $y_1 = \pi[N-1-j]$ (here we do not yet know whether $J[y_1] = I[y_1]$); the second has $x_2 = \pi[j]$ and $y_2 = \pi[N-1]$ (here we do not yet know whether $J[x_2] = I[x_2]$).

However, as we now show, because $Q_I \subset D_J$, we are able to infer that indeed $J[\pi[j]] = I[\pi[j]]$ and $J[\pi[N-1-j]] = I[\pi[N-1-j]]$. Specifically, the crucial information is provided by the values $D_J(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]])$ and $D_J(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]])$.

For any dipole $\xi = (d, \alpha, \beta)$, it will be convenient to let $\mathrm{IndCnt}(\xi)$ be the number of pairs x, y of pixels in X such that $0 < \lambda[x] < j$, $y = x + d$, $J[x] = \alpha$ and $J[y] = \beta$.

Because $Q_I \subset D_J$, we know that $D_J(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]]) = D_I(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]])$ and $D_J(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]]) = D_I(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]])$.

If the two dipoles $(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]])$ and $(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]])$ are identically equal to a single dipole ξ, then the set H_I is constructed to contain only one dipole with displacement $\pi[N-1-j]$, namely, ξ. Otherwise, H_I will contain two dipoles, ξ_1 and ξ_2, with displacement $\pi[N-1-j]$, one of which is identical to $(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]])$ and the other to $(\pi[N-1-j], I[\pi[j]], I[\pi[N-1]])$. Note that in this case we must have either $I[\pi[0]] \neq I[\pi[j]]$ or else $I[\pi[N-1-j]] \neq I[N-1]$.

If H_I contains only one dipole, ξ, with displacement $\pi[N-1-j]$, then we must have $(\pi[N-1-j], I[\pi[0]], I[\pi[N-1-j]]) = (\pi[N-1-j], I[\pi[j]], I[\pi[N-1]]) = \xi$, implying that $Q_I(\xi)$ is greater by 2 than $\mathrm{IndCnt}(\xi)$. Similarly, if H_I contains two distinct dipoles ξ_1 and ξ_2 with displacement $\pi[N-1-j]$, then for each of $i = 1$ and 2, $Q_I(\xi_i)$ must be greater by 1 than $\mathrm{IndCnt}(\xi_i)$.

We now proceed to show that $J[\pi[N-1-j]] = I[\pi[N-1-j]]$ and $J[\pi[j]] = I[\pi[j]]$. Suppose first that H_I contains two distinct dipoles with displacement $\pi[N-1-j]$: $\xi_1 = (\pi[N-1-j], \alpha_1, \beta_1)$ and

$\xi_2 = (\pi[N-1-j], \alpha_2, \beta_2)$. In this case, as observed above, for each of $i = 1$ and 2, $D_J(\xi_i) = Q_I(\xi_i)$ is greater by one than IndCnt(ξ_i). We thus infer that the dipole of J bridging pixel pair $(\pi[0], \pi[N-1-j])$ is equal to one of ξ_1 or ξ_2 and the dipole of J bridging $(\pi[j], \pi[N-1])$ is equal to the other. Since $\xi_1 \neq \xi_2$, we infer either that $\alpha_1 \neq \alpha_2$ or else that $\beta_1 \neq \beta_2$. We also know that one of α_1 or α_2 is equal to $I[\pi[0]]$ and one of β_1 or β_2 is equal to $I[\pi[N-1]]$. Suppose that $I[\pi[0]] = \alpha_1 \neq \alpha_2$. Then we infer that $J[\pi[j]] = \alpha_2$ and $J[\pi[N-1-j]] = \beta_1$; however, the construction of H_I requires that $\alpha_2 = I[\pi[j]]$ and $\beta_1 = I[\pi[N-1-j]]$, implying that $J[\pi[j]] = I[\pi[j]]$ and $J[\pi[N-1-j]] = I[\pi[N-1-j]]$. Similarly, in the case in which $I[\pi[N-1]] = \beta_1 \neq \beta_2$, we find that $J[\pi[j]] = I[\pi[j]]$ and $J[\pi[N-1-j]] = I[\pi[N-1-j]]$.

If H_I contains only a single dipole $\xi = [\pi[N-1-j], \alpha, \beta]$, then (as argued above) Q_I's construction ensures that $Q_I(\xi)$ is greater by two than IndCnt(ξ). Since the only pixel pairs with displacement $\pi[N-1-j]$ that remain unaccounted for are $(\pi[0], \pi[N-1-j])$ and $(\pi[j], \pi[N-1])$, we infer that $\chi_1 = \chi_{j+1} = \xi$, implying that $J[\pi[j]] = J[\pi[0]] = \alpha = I[\pi[j]] = I[\pi[0]]$ and $J[\pi[N-1-j]] = J[\pi[N-1]] = \beta = I[\pi[N-1-j]] = I[\pi[N-1]]$.

This completes the proof in the case in which N is assumed to be even. No important complexities arise if N is odd. ■

5. FINAL REMARKS

Visual processes such as recognition of an object or face are concerned primarily with the spatial relations between intensities composing the input image. The dipole histogram of an image encodes purely relational information about that image. It is also true that the dipole histogram of any discrete, finite image uniquely determines that image.[23] These considerations make the dipole histogram a promising way of representing an image, at least for purposes of recognition.

However, it does not seem likely that human vision encodes its input images in terms of their dipole histograms. First, the definition of a dipole histogram implicitly requires an unrealistic level of precision in registration of image intensity, point by point throughout the stimulus field. Second, the dipole histogram of a given image is typically of much larger cardinality than the image itself.

In this paper we have attempted to address these two difficulties in order to explore the possibility that some form of purely relational image representation might subserve recognition processes in human vision. In Section 3 we introduced soft dipole representations. These representations are similar to dipole histograms; however, they forgo the precise registration of image intensity required by the dipole histogram. We also showed that soft dipole representations uniquely determine the images to which they correspond.

It should be noted, however, that the soft dipole representations we have considered here tend to be large. An important open question (raised in Subsection 3.C) is the following: Do there exist *small soft* dipole representations that suffice to uniquely determine the images to which they correspond?

That there do, in fact, exist small dipole representations was demonstrated in Section 4. In that section we provided a method for constructing a purely relational representation Q_I of any discrete, finite image I with $N > 1$ pixels, such that $|Q_I| \leqslant N-1$. Given the information that (a) I has N pixels and (b) Q_I is a subset of D_I (the dipole histogram of I), one can uniquely determine the pixel values of I. This result demonstrates that one can always obtain a purely relational representation of a given image that is no greater in order of complexity than the original image.

It should be noted, however, that the construction we have provided does not guarantee that the resulting relational representation will be minimal in cardinality. There may well exist smaller subsets of D_I that uniquely determine I. For example, as discussed in Subsection 4.A, in the special case in which all of I's pixels take the same value, I is uniquely determined by a restriction $Q \subset D_I$ for which $|Q| = 1$ (as usual, provided that one knows the number of pixels in I). The method we have described yields representations of cardinality at least ceil($N/2$), for ceil($N/2$) the greatest integer less than or equal to $N/2$.

Thus the following problem presents itself: Given an arbitrary input image I, how can one find a dipole representation of I of minimal cardinality? That is, is there a general method for constructing a restriction Q of D_I such that (1) Q uniquely determines I and (2) $|Q|$ is minimal over the set of all restrictions satisfying (1)?

We speculate that a solution to this problem may yield interesting insights into the nature of visual coding. Doner[24] investigated the relationship between perceptual properties of images I and the corresponding dipole histograms D_I. In particular, he examined the ways in which various measures of dipole histogram entropy relate to different sorts of pattern judgments. Although this idea is enticing, we suspect that it may be important to focus on the relationships between visual judgments and *nonredundant* dipole representations.

To help clarify our intuitions we offer the following definition: For any discrete, finite image I, define the *dipole complexity* of I as the minimum of $|Q|$ across all restrictions Q of D_I that uniquely determine I.

What might the visual significance be (if any) of the dipole complexity of an image? We have already observed that the dipole complexity of a one-dimensional image of uniform color is 1. More generally, it seems reasonable to suspect that images of low dipole complexity may turn out to be structurally "simple" in some sense echoed by human pattern perception, whereas those of high dipole complexity may turn out to be visually "complex."

The current paper shows that the dipole complexity of any image I is less than or equal to $|I| - 1$. At present, however, the problem of how to determine the dipole complexity of a given image remains unsolved.

APPENDIX A

In this appendix, for the sake of building intuition, we supply a proof, previously presented,[23] that any one-dimensional image is uniquely determined by its dipole histogram. This is, of course, a corollary of Proposition 4.C.2. However, the following proof is simpler than that of Proposition 4.C.2.

Proof. Let I be a one-dimensional image. It will be convenient to write $\Sigma_{\alpha,\beta}$ to indicate a sum ranging over all pairs of pixel values of I. We use this notation to define

$$C_{11}[d] = \sum_{\alpha,\beta} D_I[d, \alpha, \beta]\alpha\beta = \sum_{r=0}^{N-1-d} I[r]I[r+d] \tag{A1}$$

and

$$C_{10}[d] = \sum_{\alpha,\beta} D_I[d, \alpha, \beta]\alpha(1-\beta) = \sum_{r=0}^{N-1-d} I[r](1 - I[r+d]). \tag{A2}$$

Now, for $k = 0, 1, \ldots, N-1$, we have

$$C_{11}[N-1-k] + C_{10}[N-1-k] = \sum_{r=0}^{k} I[r]I[r+N-k] + \sum_{r=0}^{k} I[r](1 - I[r+N-k]) = \sum_{r=0}^{k} I[r]. \tag{A3}$$

Thus, immediately,

$$I[0] = C_{11}[N-1] + C_{10}[N-1], \tag{A4}$$

and with $I[0]$ in hand, we recursively obtain $I[k]$ for $k = 1, 2, \ldots, N-1$ as follows:

$$I[k] = C_{11}[N-1-k] + C_{10}[N-1-k] - \sum_{r=0}^{k-1} I[r]. \tag{A5}$$

Address correspondence to Charles Chubb at the location on the title page or by e-mail, cfchubb@uci.edu; phone, 949-824-1481; or fax, 949-824-2517.

REFERENCES

1. R. Desimone and C. G. Gross, "Visual areas in the temporal cortex of the macaque," J. Neurosci. **4**, 2051–2062 (1979).
2. M. Ito, H. Tamura, I. Fujita, and K. Tanaka, "Size and position invariance of neuronal responses in monkey inferotemporal cortex," J. Neurophysiol. **73**, 218–226 (1995).
3. E. Kobatake and K. Tanaka, "Neural selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex," J. Neurophysiol. **71**, 856–867 (1994).
4. K. Tanaka, H. Saito, Y. Fukada, and M. Moriya, "Coding visual images of objects in the inferotemporal cortex of the macaque monkey," J. Neurophysiol. **66**, 170–189 (1991).
5. D. I. Perrett, E. T. Rolls, and W. Caan, "Visual neurons responsive to faces in monkey temporal cortex," Exp. Brain Res. **47**, 329–342 (1982).
6. E. T. Rolls and G. C. Baylis, "Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey," Exp. Brain Res. **65**, 38–48 (1986).
7. P. Diaconis and D. Freedman, "On the statistics of vision: the Julesz conjecture," J. Math. Psychol. **24**, 112–138 (1981).
8. A. Gagalowicz, "A new method for texture fields synthesis: some applications to the study of human vision," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-3**, 520–533 (1981).
9. E. N. Gilbert, "Random colorings of a lattice on squares in the plane," SIAM J. Alg. Disc. Meth. **1**, 152–159 (1980).
10. B. Julesz, "Visual pattern discrimination," IRE Trans. Inf. Theory **IT-8**, 84–92 (1962).
11. B. Julesz, "Textons, the elements of texture perception, and their interactions," Nature **290**, 91–97 (1981).
12. B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second order statistics," Perception **2**, 391–405 (1973).
13. B. Julesz, E. N. Gilbert, and J. D. Victor, "Visual discrimination of textures with identical third order statistics," Biol. Cybern. **31**, 137–149 (1978).
14. J. D. Victor, "Images, statistics and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: comment," J. Opt. Soc. Am. A **11**, 1680–1684 (1994).
15. J. D. Victor and S. Brodie, "Discriminable textures with identical Buffon needle statistics," Biol. Cybern. **31**, 231–234 (1978).
16. J. D. Victor, M. Conte, K. Purpura, and E. Katz, "Isodipole textures: a window on cortical mechanisms of form processing," in *Early Vision and Beyond*, T. V. Papathomas, C. Chubb, A. Gorea, and E. Kowler, eds. (MIT Press, Cambridge, Mass., 1995), pp. 99–107.
17. J. I. Yellott, "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," J. Opt. Soc. Am. A **10**, 777–793 (1993).
18. J. R. Bergen and M. S. Landy, "The computational modeling of visual texture segregation," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds., (MIT Press, Cambridge, Mass., 1991), pp. 253–271.
19. C. Chubb, J. Econopouly, and M. S. Landy, "Histogram contrast analysis and the visual segregation of IID textures," J. Opt. Soc. Am. A **11**, 2350–2374 (1994).
20. N. Graham, "Non-linearities in texture segregation," in *Higher-order Processing in the Visual System. Ciba Foundation Symposium*, Vol. 184, J. A. Goode and G. R. Bock, eds. (Wiley, Chichester, UK, 1994), pp. 309–329.
21. M. S. Landy and J. R. Bergen, "Texture segregation and orientation gradient," Vision Res. **31**, 679–691 (1991).
22. J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," J. Opt. Soc. Am. A **7**, 923–932 (1990).
23. C. Chubb and J. I. Yellott, "Every discrete, finite image is uniquely determined by its dipole histogram," Vision Res. **40**, 485–492 (2000).
24. J. Doner, "Dipole information complementarity in discrete 2D patterns," J. Math. Psychol. **43**, 355–393 (1999).