Learning by Assertion: Two Methods for Calibrating a Linear Visual System

Laurence T. Maloney
Center for Neural Science, Department of Psychology, New York University, New York, NY 10003 USA

Albert J. Ahumada
NASA Ames Research Center, Moffett Field, CA 94035 USA

A visual system is geometrically calibrated if its estimates of the spatial properties of a scene are accurate: straight lines are judged straight, angles are correctly estimated, and collinear line segments are perceived to fall on a common line. This paper describes two new calibration methods for a model visual system whose photoreceptors are initially at unknown locations. The methods can also compensate for optical distortions that are equivalent to remapping of receptor locations (e.g., spherical aberration). The methods work by comparing visual input across eye/head movements; they require no explicit feedback and no knowledge about the particular contents of a scene. This work has implications for development and calibration in biological visual systems.

1 Introduction

It's likely that no biological visual system is ever perfectly calibrated, but considerable evidence exists that biological visual systems do compensate for optical distortions and initial uncertainty about the position of photoreceptors in the retinal photoreceptor lattice (Banks 1976; Hirsch and Miller 1987). Recent anatomical work, for example, demonstrates apparent disorder in the retinal lattice outside the central fovea, increasing with eccentricity (Hirsch and Miller 1987). Further, the optics of the eye change throughout the life span (Banks 1976; Weale 1982), suggesting that calibration may continue in the adult.

Previous work in visual neural development suggests a variety of sources of information that drive calibration (Meyer 1988; Purves and Lichtman 1985; Shatz 1988), and there are computational models of visual neural development based on these cues (Sejnowski 1987). Yet, although biological visual systems are known to require patterned visual stimulation to achieve normal organization (Movshon and Van Sluyters...
1981), few models require such stimulation to function. Exceptions in-
clude Kohonen (1982) and Toet, Blom, and Koenderink (1987). Further,
while all these models could in principle compensate for disorder in the
retinal lattice, none of them addresses the problem of compensation for
optical distortion.

We describe two methods for calibrating a simple linear visual system
that work by comparing visual input across eye/head movements. These
methods can organize the receptive fields of a simple visual system so as
to compensate for irregularities in the retinal photoreceptor lattice and
optical irregularities equivalent to distortions in the lattice. They require
no explicit feedback and no knowledge about the particular contents of a
scene, but instead work by “asserting” that the internal representation of
the scene behave in a prespecified way under eye and head movements.
We demonstrate that these methods can be used to calibrate the simple,
linear visual system described next. In the final section, we discuss the
implications of this work for other models of visual processing.

2 A Model Linear Visual System

The model visual system has \( N \) photoreceptors arranged in a receptor
array. The locations of these receptors are initially unknown. The light
image is the mean intensity of light at each location in the receptor array.
The output of a receptor is the value of the light image at the location
assigned to the receptor. The mapping from light image to receptor array
is assumed to be linear. The output (measured intensity) from the \( i \)th
receptor is denoted \( \rho_i \). In vector notation, the instantaneous input from
the receptor array is \( \rho = [\rho_1, \ldots, \rho_N]^T \).

Figure 1 also shows an ideal receptor array of \( N \) receptors at specified,
known locations. This receptor array may be a square or hexagonal grid
of receptors, but it need not be. The input from the \( i \)th receptor in this
ideal array is denoted \( \mu_i \), and the input from the ideal array is denoted
\( \mu = [\mu_1, \ldots, \mu_N]^T \).

The real array is connected to the ideal by linear receptive fields (one is
shown in the figure). The visual system is calibrated when the receptive
fields translate the input of the irregular, real array to what the ideal
array would have sampled.

Without some restriction on the light images sampled, there need be,
of course, no connection between samples taken by the real array and
those taken by the ideal array. For the remainder of this paper, the set of
light images \( L \) is assumed to be a space of two-dimensional finite Fourier
series of dimension \( N \) where \( N \) is the number of receptors in the ideal
and in the real arrays. When \( N \) is 49, as in the simulations below, the 49-
dimensional space of two-dimensional finite Fourier series are weighted
sums of products of one of

\[
1. \sin 2\pi x \cdot \cos 2\pi x \cdot \sin 2\pi 2x \cdot \cos 2\pi 2x \cdot \sin 2\pi 3x \cdot \cos 2\pi 3x
\]
and one of

$$1, \sin 2\pi y, \cos 2\pi y, \sin 2\pi 2y, \cos 2\pi 2y, \sin 2\pi 3y, \cos 2\pi 3y$$

This lowpass assumption reflects the blurring induced by the optics of the eye; it is commonly made in modeling spatial vision (Maloney 1990). With this assumption, we can show that if there is a solution to the calibration problem for a particular real array, ideal array, and linear subspace of lights, then it must be a linear transformation (Maloney 1990). This linear transformation, $W$, will depend on the unknown position of receptors in the real array and, consequently, on the optics of the visual process.
system. Each row of the linear transformation $W$, written as a matrix, is the weights of the receptive field of one ideal receptor.

Because of the irregular distribution of receptors in the real array, the receptive fields for different receptors in the ideal array may be very different. Figure 2 shows two correctly calibrated receptive fields corresponding to two locations in the ideal array pointed to by dashed lines. The receptors in the real array are shown as black dots. The weights assigned to the nearby receptors in the real array are shown in the "exploded" squares at the ends of the dashed lines. These two receptive fields both extract the equivalent information that would have been sampled by an ideal receptor at the location indicated.

The problem of calibration is now reduced to learning the unknown linear transformation $W$, given input only from the real sampling array.

Figure 2: Equivalent receptive fields. The black dots represent the locations of photoreceptors in the disordered real array. The squares correspond to the receptive fields of two receptors in the ideal array. See text.
In computer vision this problem is commonly solved by the use of "test patterns." If the contents of the scene are known, then the value of the ideal array, \( p \), can be compared to the value that is correct for the known test pattern, and \( W \) can be adjusted to eliminate any discrepancies (Rosenfeld and Kak 1976). We describe algorithms that do not require knowledge of the specific contents of the scene.

3 Eye/Head Movements

Consider the consequences of moving the eye and/or head, while the scene remains unchanged. The eye and head may translate to a new position, change the angle of gaze, rotate, "zoom" in or out on the scene, and so on. A particular eye and head movement serves to transform the value of the real (retinal) array (see Fig. 3). The transform, denoted \( T \), maps the initial value of the real array \( \rho \) to the value after an eye/head movement, \( \rho' \). Different eye/head movements, of course, correspond to different transformations, \( T \). If the visual system is properly calibrated, then \( \mu = W\rho \) and \( \mu' = W\rho' \) will be related by an equivalent transformation, denoted \( t \). Intuitively, if the retinal image moves rigidly to the left on the real array, then, in a calibrated visual system, it would move rigidly to the left on the ideal array as well.

\( T \) is a physical transformation induced by actual eye/head movements. \( t \) is an internal transformation that simulates the external transformation. The last assumption we make concerning the model visual system is that it can perform transformations, denoted \( t \), on the ideal array that mimic all possible eye and head movements. The set of transfor-

![Figure 3: Schematic diagram of an assertion. The consequences of an eye/head movement can be computed in two ways. See text for an explanation.](http://www.mitpressjournals.org/doi/pdfplus/10.1162/neco.1989.1.3.392)
mations $t$ is easily computed; it is precisely the transformations needed, for example, to compensate computationally for eye movements.

The visual system can now compute the outcome of eye/head movements in two ways. It can look at the scene, take the resulting value $\mu = W\rho$ and apply $t$ to get $\mu'$. Alternatively, it can perform the physical transformation $T$ by actually moving, and then compute $\mu' = W'T\rho$. If the two methods of computing $\mu' = tW\rho = W'T\rho$ (the two paths sketched by arrows in Fig. 3) produce different answers, then the visual system is not calibrated. Conversely, a specific transformation $T$ constrains the choice of $W$ so that $tW = W'T$. This constraint we term an assertion. We assert, for example, that in a calibrated visual system, moving closer to an object should simply result in scaling the object in size. Any other changes (rippling, flickering, distortion) are indications of failures of calibration.

4 Mathematical Results

To what extent is $W$ constrained by all of the transformations $T$ taken together? For the simple visual system considered here, we have the following mathematical results (Maloney 1989):

**Result 1.** If $W$ is nonsingular, it is completely determined up to a scaling factor by the assertions generated by all eye and head movements in the scene.

Satisfying the assertions is almost equivalent to calibrating the visual system. The requirement that $W$ be nonsingular avoids pathological solutions where the visual system disconnects itself from the environment. If, for example, all weights in $W$ were set to 0, the visual system, missing all visual input, could never see any failure of rigid transformation under eye/head movements.

The second result concerns equivalent receptive fields. Suppose we consider not all eye and head movements, but just (small) eye movements: translations of the retina perpendicular to the line of sight.

**Result 2.** If $W$ is nonsingular, and any one receptive field of the ideal array is fixed, then all the other receptive fields of the ideal array are completely determined by the assertions generated by eye movements that translate the retinal image.

The importance of the second result is that it would permit anatomically more regular portions of the retina to serve as a template for organizing equivalent receptive fields elsewhere in the retina.

The two results, taken together, suggest that assertions can be used to guide calibration. In the next sections, we develop and illustrate an algorithm for calibration based on assertions.
5 Learning by Assertion

The requirement that $WT = tW$ for all transformations $T$ and all light images determines $W$ as stated above. The penalty term

$$\text{Pen}(W) = \sum_T \sum_\rho |WT(\rho) - tW(\rho)|^2$$  \hspace{1cm} (5.1)

is minimized precisely when this condition holds. We can therefore develop a learning algorithm for calibrating the simple visual system outlined above by minimizing the quadratic penalty in equation 5.1.

The algorithm repeats the following steps until the error term (equation 5.1) is sufficiently small:

1. Generate a light image drawn at random from the lowpass subspace of finite Fourier series.
2. Sample the light image to obtain $\rho$.
3. Simulate a randomly chosen eye/head movement $T$.
4. Resample the light image to obtain $\rho'$.
5. Compute $WT(\rho) - tW(\rho)$, the Euclidean vector difference between the two ways to compute $\rho'$.

The Widrow-Hoff algorithm (Widrow and Hoff 1960) compares the correct and actual outputs of a linear transformation and alters the transformation to make them coincide. Our algorithm is computationally identical to the Widrow-Hoff algorithm except that it compares $tW\rho$ and $WT\rho$, two possibly erroneous outputs, and attempts to minimize the discrepancy between them. We do not know the conditions under which the modified algorithm is guaranteed to converge.

In Step 5, we assume that the transformation $t$, corresponding to $T$, is known to the visual system. We are investigating whether $t$ can be estimated from the visual input with inexact knowledge of $T$ (Maloney 1989).

6 Simulation Results

We have implemented the algorithm corresponding to Result 2 above: eye movements only. For the results reported here we assumed a $7 \times 7$ square grid ideal array with a 49-receptor irregular real array. The locations of the 49 receptors in the real array were chosen by randomly perturbing about half of the receptors in the regular array by about 0.25 of the spacing between receptors in the regular array. The lowpass space was described above. One receptive field in the ideal array was fixed.
Figure 4a shows one of the $7 \times 7$ learned receptive fields after 150,000 iterations of the Widrow–Hoff algorithm; Figure 4b shows the correct receptive field for this element.

A linear interpolation algorithm was used to render each $7 \times 7$ grids as a perspective plot. The receptive field has converged to its desired shape.

**7 Generalizations and Predictions**

We are currently implementing the algorithm corresponding to Result 1 (eye and head movements). For a nonlinear visual system, equation 5.1 still serves as a constraint guiding calibration that in combination with other constraints may be sufficient to guarantee proper calibration. Since equation 5.1 is quadratic, it is plausible that any candidate neural learning algorithm is capable of minimizing it, if the penalty term can be
computed. Since the penalty term represents nonrigid motion induced in the representation by eye or head movements, it is plausible that the penalty term is available to the nervous system.

The methods described here use a novel cue, derived from comparison of visual input across successive glances at a scene, to calibrate a simple linear visual system. Previous models of visual neural development, reviewed above, use different cues and methods to calibrate the visual system. The methods outlined here differ from these methods in that (1) they directly optimize a visual capability (stability and rigidity under change of direction of gaze), (2) they can compensate for small optical distortions and remappings, (3) they require structured visual input (actual scenes), and (4) they require successive fixations on a single unchanging scene.

Taken as a claim about visual development, the procedures developed here are readily testable empirically. (1) Animals reared in environments lacking structured visual input, or in environments where visual input is rendered perpetually nonrigid, or where it is never possible to fixate the same scene twice, should be perfectly calibrated according to previous theories, but not according to the work developed here. (2) Animals with small optical distortions induced early in development should not be perfectly calibrated according to previous theories, but may be so according to the work described here. (3) The visual system will compensate for retinally stabilized optical distortions in adults; these distortions may include small induced scotomas. Prediction (2) may hold while (3) fails if recalibration in the adult is limited by connectivity restrictions on receptive fields (Purves and Lichtman 1985).

Acknowledgments

We thank Marty S. Banks, Jeffrey B. Mulligan, Michael Pavel, Brian A. Wandell, and John I. Yellott, Jr. for useful comments. Send reprint requests to Laurence T. Maloney, Department of Psychology, Center for Neural Science, 6 Washington Place, 8th Floor, New York, NY 10003.

References


Received 23 May 1989; accepted 13 July 1989.