An LGN Inspired Detect/Transmit Framework for High Fidelity Relay of Visual Information with Limited Bandwidth

Nicholas A. Lesica and Garrett B. Stanley

Division of Engineering and Applied Sciences, Harvard University,
Cambridge MA 02138, USA
lesica@fas.harvard.edu

Abstract. The mammalian visual system has developed complex strategies to optimize the allocation of its limited attentional resources for the relay of behaviorally relevant visual information. Here, we describe a framework for the relay of visual information that is based on the tonic and burst properties of the LGN. The framework consists of a multi-sensor transmitter and receiver that are connected by a channel with limited total bandwidth. Each sensor in the transmitter has two states, tonic and burst, and the current state depends on the salience of the recent visual input. In burst mode, a sensor transmits only one bit of information corresponding to the absence or presence of a salient stimulus, while in tonic mode, a sensor attempts to faithfully relay the input with as many bits as are available. By comparing video reconstructed from the signals of detect/transmit sensors with that reconstructed from the signals of transmit only sensors, we demonstrate that the detect/transmit framework can significantly improve relay by dynamically allocating bandwidth to the most salient areas of the visual field.

1 Introduction

The mammalian early visual pathway serves to relay information about the external world to higher brain areas where it can be analyzed to make decisions and govern behavior. However, this relay is constrained by the availability of limited attentional resources. Because mammals can only attend to a small fraction of the visual field at any given time, the early visual pathway must carry out two distinct tasks: the detection of salient input to direct the deployment of attentional resources and transmission of detailed features of those stimuli to higher brain areas. Neurons in the lateral geniculate nucleus (LGN) of the thalamus have two response modes known as tonic and burst, and there is evidence that these response modes serve to facilitate the tasks of detection and transmission (for review, see [1,2]).

The LGN relays the output of the visual system's peripheral sensors in the retina, making both feedforward and feedback connections with the visual system's computational center in the cortex. The response mode of an LGN neuron is determined by the state of a special set of low-threshold voltage-dependent channels

M. De Gregorio et al. (Eds.): BV&AI 2005, LNCS 3704, pp. 177–186, 2005.

known as T channels [3]. When the membrane is depolarized and the neuron is firing frequently, the T channels are inactivated, and the neuron is in tonic mode. In tonic mode, the spontaneous firing rate is high, and modulations in the response are linearly related to modulations in the visual input, allowing the neuron to faithfully relay both excitatory and inhibitory features to the cortex. When the membrane is hyperpolarized for a prolonged period of time and the neuron is silent, the T channels are de-inactivated and the neuron enters burst mode. When the neuron is in burst mode, depolarization of the membrane opens the T channels, resulting in a wave of current which further depolarizes the membrane and causes a stereotyped burst of closely spaced action potentials. This allows the neuron to signal the appearance of a input with an amplified response.

During visual stimulation, the membrane potential (and thus, response mode) of an LGN neuron is controlled in part by feedback connections from the cortex [4]. Thus, the thalamocortical circuit is thought to perform both detection and transmission as follows: In the absence of a salient input, the membrane is hyperpolarized, the T channels are de-inactivated, and the neuron is in burst mode. Upon the appearance of a salient stimulus, the membrane is briefly depolarized and a burst is triggered. Cortical feedback then maintains the depolarization of the neuron, switching it to tonic mode and increasing the spontaneous firing rate. While the stimulus persists, tonic firing transmits detailed information about the stimulus. When the stimulus disappears, the neuron falls silent, cortical feedback hyperpolarizes the membrane, and the cycle repeats. This silence/burst/tonic/repeat response pattern has been observed in both anesthetized and awake animals, in the LGN responses to sinusoidal gratings [5,6] and natural scene movies, as objects moved in and out of the receptive field [7].

Here, we develop a detect/transmit framework for the relay of visual information based on the tonic and burst properties of the LGN. The framework consists of a multi-sensor transmitter (LGN) and receiver (cortex) that are connected by a channel with limited total bandwidth (attention). Each sensor in the transmitter has two states: tonic and burst. In burst mode, a sensor transmits only one bit of information corresponding to the absence or presence of a salient stimulus. In tonic mode, a sensor attempts to faithfully relay the visual input with as many bits as are available. The mode of each sensor is determined by the salience of the recent visual input. To evaluate the detect/transmit framework, we compare video reconstructed from the outputs of detect/transmit sensors with that reconstructed from the outputs of transmit only sensors. The results demonstrate that the detect/transmit framework can significantly increase the fidelity of relay by dynamically allocating bandwidth to the most salient areas of the visual field.

2 A Detect/Transmit Framework for the Relay of Visual Information

Based on the tonic and burst properties of the LGN that facilitate the detection and transmission of visual inputs, we have developed a framework for the high

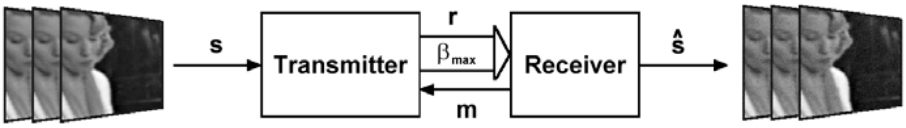

Fig. 1. An LGN inspired scheme for the relay of visual information

fidelity relay of visual information over a channel with limited bandwidth. The framework consists of a multi-sensor transmitter with tonic and burst modes, and a receiver that decodes the transmitted signal and controls the mode of each sensor in the transmitter, designed to mimic cortical feedback control of LGN response mode. A schematic diagram of the framework is shown in figure 1.

The intensity of the visual stimulus (s) is specified by P pixels per frame. The transmitter contains P sensors, each of which corresponds directly to one pixel of the visual input. The transmitter sends the output of each sensor to the receiver once per frame via a noise-free, lossless channel. The bandwidth limit on the channel (for all sensors combined) is specified as β_{max} bits/sec, which, for a frame rate of F frames/sec, corresponds to $\beta_{max}/F = \beta_{frame}$ bits/frame. Each sensor in the transmitter can operate in either tonic or burst mode. In tonic mode, the sensor will attempt to transmit detailed features of the visual stimulus with as many bits as are available. In burst mode, the sensor will signal either the absence or presence of a salient stimulus with only one bit. Following the relay of each frame, the receiver determines the mode (m) of each sensor for the next frame based on the salience of the recent visual input and sends the modes back to the transmitter (Note that the P bits/frame required to send the mode signal back to the transmitter is additional and is not included in constraint β_{max}).

We designed the detect/transmit framework to mimic the ability of the mammalian visual system to efficiently transmit visual information based on 'bottom-up' control of attention in response to changes in the external environment. However, 'bottom-up' control of attention is only one of many strategies that the visual system has developed to improve the transmission of visual information. Other strategies, such as spatial and temporal decorrelation, separate ON and OFF channels, and mechanisms for task dependent 'top-down' control of attention are not included in the model. Correspondingly, in evaluating the framework, we assumed that the goal of the transmitter is to send a representation of the visual stimulus with minimal mean-squared error (MSE). Thus, our model neglects any other features of the neural response that may be important, such as sparseness or redundancy [8]. Further discussion can be found in section 4.

2.1 Transmitter

The operation of the transmitter can be divided into three steps that must be repeated for each frame of the input. First, the total bandwidth β_{frame} is distributed among the P sensors in the transmitter based on the mode signal m

sent back from the receiver. Next, the recent history of the input is evaluated for comparison with the current input. Finally, the output of each sensor is calculated and sent to the receiver. Each of these steps is described in detail below.

Distribute bandwidth: Let the number of sensors in burst and tonic mode at a given time as determined by the mode signal m be denoted by n_{burst} and n_{tonic}. The total bandwidth β_{frame} must be distributed among the P sensors based on their modes. Each burst sensor is allotted one bit ($\beta_{burst} = 1$), and the remaining bits are distributed among the tonic sensors as follows:

$$\beta_{tonic} = floor\left\{\frac{\beta_{frame} - n_{burst}}{n_{tonic}}\right\}$$

Thus, at time step t, the number of bits available to a given sensor, $\beta(p,t)$, is determined based on its mode $m(p,t)$ as follows:

$$\beta(p,t) = \begin{cases} \beta_{burst} & , \quad m(p,t) = 0 \\ \beta_{tonic} & , \quad m(p,t) = 1 \end{cases}$$

Evaluate input history: For each sensor, the recent input history must be evaluated to determine the salience of the current input. Typically, the salience of the input in a particular region of the visual field is evaluated across multiple dimensions (orientation, color, contrast, etc.) [9]. Here, salience is measured independently for each pixel by simply comparing the current intensity to previous intensities.

For a given sensor, the recent history of the input, H_{burst}, is specified by the average of the previous α intensities of the corresponding pixel:

$$H_{burst}(p,t) = \sum_{k=1}^{\alpha} s(p,t-k)$$

where α specifies the number of frames to be considered in the history of the input. If a sensor is in burst mode, it will signal a change in the input if the current input is significantly different from H_{burst}, alerting the receiver to switch the sensor to tonic mode. For all sensors that have just switched from tonic to burst mode at time t ($m(p,t-1) = 1$ and $m(p,t) = 0$), the history term H_{burst} must be updated. For all sensors that remain in burst mode from the previous time step ($m(p,t-1) = 0$ and $m(p,t) = 0$), H_{burst} remains the same ($H_{burst}(p,t) = H_{burst}(p,t-1)$). For all sensors in tonic mode, the input history is evaluated at the receiver as described below.

Send signal: Once the mode, available bandwidth, and recent input history for each sensor have been set, the transmitter can relay its output to the receiver. The output of a sensor in burst mode depends on the salience of the current input relative to the recent history H_{burst}, with sensitivity determined by the parameter σ_{burst}. If the current input is significantly different from the recent history, then the sensor will indicate a change:

$$r(p,t) = \begin{cases} 1 & , \quad |s(p,t) - H_{burst}(p,t)| \geq \sigma_{burst} \\ 0 & , \quad \text{otherwise} \end{cases}$$

Sensors in tonic mode simply relay the visual input, quantized to available number of bits β_{tonic}:

$$r(p,t) = Q(s(p,t), \beta_{tonic})$$

where Q is the quantizer function.

2.2 Receiver

Receive signal: For sensors in burst mode, the receiver assumes that the input is unchanged, regardless of the transmitter output. Of course, if the sensor is in burst mode and the output $r(p,t) = 1$, the receiver will switch the sensor to tonic mode for the next frame (see below), but has received no new information about the input for the current frame. For sensors in tonic mode, the current value of the input has been relayed. Thus, the input s is reconstructed at the receiver as follows:

$$\hat{s}(p,t) = \begin{cases} \hat{s}(p,t-1) & , \quad m(p,t) = 0 \\ r(p,t) & , \quad m(p,t) = 1 \end{cases}$$

Evaluate input history: Just as the transmitter uses the recent input history to determine when the input changes significantly and signals the switch from burst to tonic mode, the receiver must determine when the input is no longer changing to control the switch back to burst mode. The switch from tonic to burst mode is controlled by comparing the current reconstruction to the recent history. At each time step, the history term for each tonic sensor is updated as follows:

$$H_{tonic}(p,t) = \sum_{k=1}^{\alpha} \hat{s}(p,t-k)$$

where α specifies the number of frames to be considered in the history of the reconstructed input.

Set mode: For each sensor, the mode for the next frame is determined by the current reconstruction and its recent history. Burst sensors that did not signal a change in the input at time t ($r(p,t) = 0$) remain in burst mode, while those that did ($r(p,t) = 1$) switch to tonic mode. The mode of each tonic sensor is determined by comparing the current reconstruction with H_{tonic} as follows:

$$m(p,t) = \begin{cases} 1 & , \quad |\hat{s}(p,t) - H_{tonic}| \geq \sigma_{tonic} \\ 0 & , \quad \text{otherwise} \end{cases}$$

The modes are sent back to the transmitter and the process is repeated for the next frame.

3 Examples of Video Relay with the Detect/Transmit Framework

To demonstrate the performance of the detect/transmit framework, we used it to relay and reconstruct a video movie. The video that we used contains footage of a vehicle traffic intersection in Karlsruhe, Germany, taken by a stationary camera. The video was provided by the Institut für Algorithmen und Kognitive Systeme, Universität Karlsruhe (http://i21www.ira.uka.de/image_ sequences). We used a section of the video consisting of 1000 frames, each of which contains 100 × 100 8-bit (0 - 255) grayscale pixels. In addition to reconstructing video from

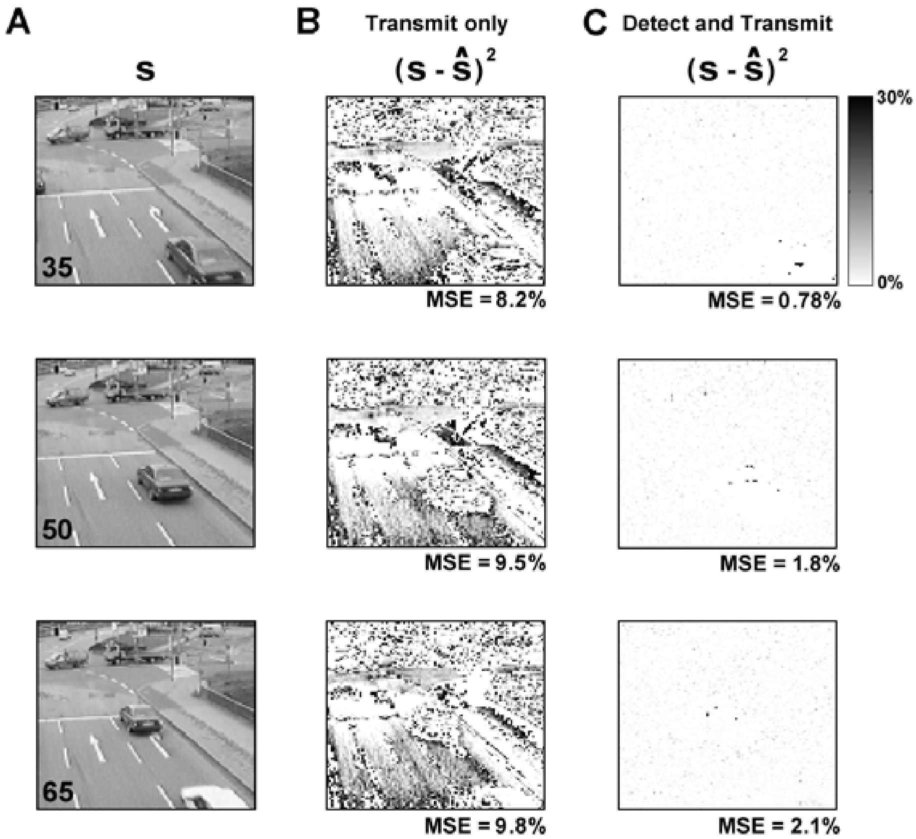

Fig. 2. Actual frames from the traffic video and the error in the reconstructions. Each frame consisted of 100 × 100 8-bit grayscale pixels. (A) Actual frames 35, 50, and 65. (B) Squared error in the reconstructed frames (% variance of intensity of actual frame) from TO sensors with $\beta_{frame}/P = 3$. The MSE of each reconstructed frame is shown. (C) Squared error in the reconstructed frames from D/T sensors with $\beta_{frame}/P = 3$ and $\sigma_{tonic} = \sigma_{burst} = 2$.

the signals of detect/transmit (D/T) sensors, we also reconstructed video from transmit only (TO) sensors as a baseline for comparison. To initialize the relay, all sensors were set to burst mode and the first α frames of the reconstructed input were set to the same value as the actual input. Because the frame rate of the video was 30 frames/sec, a value of $\alpha = 3$ was used so that the timescale of the history term was similar to the time constant of T channel de-inactivation in the LGN [3].

Example frames of the actual video and the error in the reconstructions are shown in figure 2. Figure 2A shows actual frames 35, 50, and 65 of the video. Figure 2B shows the squared error in the reconstructed frames (as a percent of the variance of the intensity of the actual frame) from relay with TO sensors

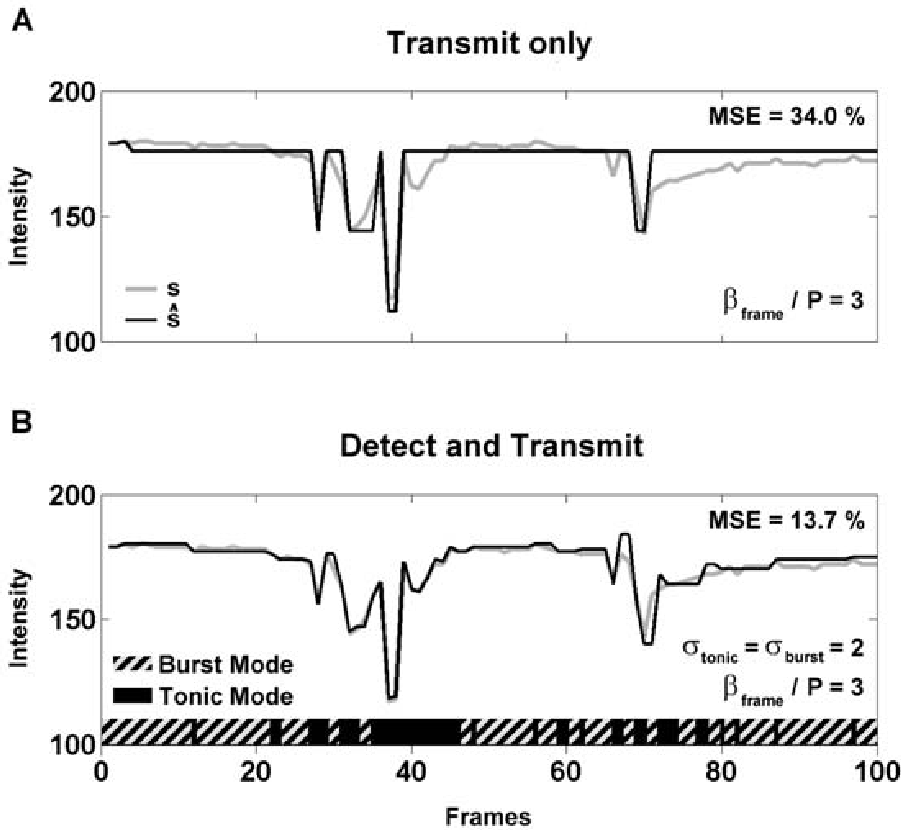

Fig. 3. Actual and reconstructed intensities of one pixel of the traffic video over 100 frames. (A) The actual (gray) and reconstructed (black) intensities from a TO sensor with $\beta_{frame}/P = 3$. The MSE of the reconstruction is shown (% variance of intensity of actual pixel). (B) The actual (gray) and reconstructed (black) intensities from a D/T sensor with $\beta_{frame}/P = 3$ and $\sigma_{tonic} = \sigma_{burst} = 2$.

with bandwidth limited to 3 bits/frame per sensor ($\beta_{frame}/P = 3$). The MSE of each reconstructed frame is also shown. Figure 2C shows the squared error in the reconstructed frames from relay with D/T sensors with $\beta_{frame}/P = 3$ and $\sigma_{tonic} = \sigma_{burst} = 2$. The reconstructions from the signals of the D/T sensors are superior to those from the TO sensors, as indicated by the decreased MSE.

Figure 3 shows actual and reconstructed intensities of one pixel of the video over 100 frames. Figure 3A shows the actual (gray) and reconstructed (black) intensities from a TO sensor with $\beta_{frame}/P = 3$, along with the corresponding MSE. Figure 3B shows the actual (gray) and reconstructed (black) intensities from a D/T sensor with $\beta_{frame}/P = 3$ and $\sigma_{tonic} = \sigma_{burst} = 2$. The mode of the sensor during the relay of each frame is indicated. During those times when the input is not changing, the sensor is in burst mode. Thus, it requires only 1 bit to transmit its signal, allowing the limited available bandwidth to be allocated to other sensors with more salient input. During those times when the input is varying, the sensor switches to tonic mode and transmits the value of the input

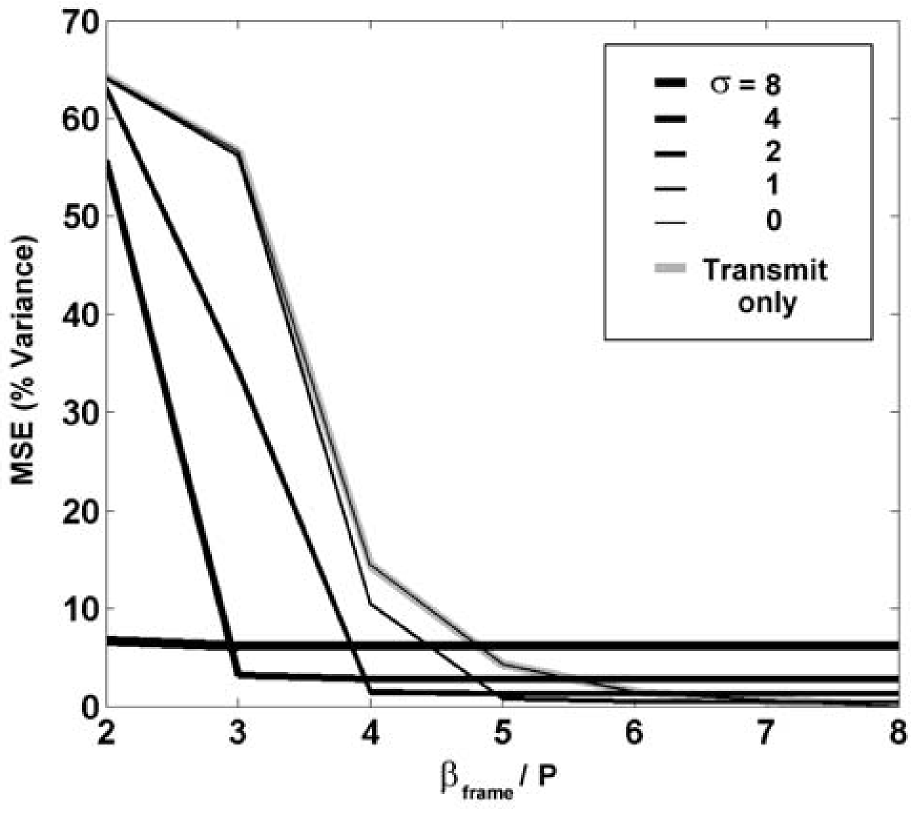

Fig. 4. Reconstruction error depends on β_{max} and σ. The MSE in the reconstruction from D/T sensors is shown for various values of β_{frame}/P and σ (see legend). For reference, the MSE in the reconstruction from TO sensors is shown in gray.

with all available bits. The dynamic allocation of bandwidth provided by the detect/transmit framework improves the reconstruction, as illustrated by the decreased MSE.

To investigate the effects of the salience sensitivity and the total available bandwidth on the fidelity of relay, we reconstructed the video from the signals of D/T and TO sensors with a range of values of β_{max} and $\sigma_{tonic} = \sigma_{burst} = \sigma$. The MSE of the reconstructions over all pixels and frames of the video are shown in figure 4. When a relatively small amount of total bandwidth is available, the lowest MSE is given by the reconstruction from the relay with the least sensitive sensors (thick black lines). This result indicates that, when bandwidth is severely limited, a better reconstruction is achieved by having fewer sensors in tonic mode with more available bits per sensor than by dividing the available bandwidth among many sensors. As more total bandwidth becomes available, the lowest MSE is given by the reconstruction from the relay with the most sensitive sensors (thin black lines). This result indicates that, when there is enough total bandwidth to encode all of the variations in the input, the best reconstruction is achieved when small fluctuations are detected.

4 Discussion

We have developed a detect/transmit framework based on the tonic and burst properties of LGN neurons to facilitate the high fidelity relay of visual information with limited bandwidth. The framework enables the dynamic allocation of bandwidth to those sensors which correspond to the most salient areas of the visual field. Each sensor in the transmitter operates in either tonic mode (signals input intensity with all available bits) or burst mode (signals the absence or presence of a salient input with only 1 bit), depending on the control signal sent by the receiver. We have demonstrated that video reconstructions from the signals of detect/transmit (D/T) sensors are superior to reconstructions from transmit only (TO) sensors and our results illustrate that the minimum MSE reconstructions are obtained when the sensitivity of the sensors (σ) is set to an appropriate value for the total available bandwidth (β_{max}).

We designed the detect/transmit framework to mimic the ability of the mammalian visual system to dynamically allocate attentional resources to behaviorally relevant areas of the visual field. However, our framework only includes mechanisms for 'bottom-up' control of attention based on changes in the external environment, and, correspondingly, control of transmitter mode was based solely on the salience of the input [10]. However, the mammalian visual system also contains mechanisms for 'top-down' control of attention that is dependent on the current behavioral task [11]. For example, if an animal is expecting something to appear in a certain area of the visual field, it may direct its attention to that area before anything actually appears. Modifications to the detect/transmit framework to incorporate 'top-down' attention would be made at the receiver, specifically to the method used to control the mode of the transmitter sensors.

In addition to attentional mechanisms, the mammalian visual system incorporates a number of other strategies to optimize the relay of visual information. While the sensors in our transmitter have a one-to-one correspondence with a pixel of the visual input, retinal ganglion cells, which transmit visual information from the retina to the LGN, are known to integrate the inputs of many photoreceptors over space and time to enhance contrast sensitivity and reduce the redundancy in their responses [12,13,14]. To incorporate these principles into our framework, each sensor would need to integrate multiple pixels of the visual input into its output and the reconstruction scheme in the receiver would have to be changed accordingly. The development of such modifications and the implementation of 'top-down' attention as described above are directions for future research.

Acknowledgments

This work was supported by National Geospatial-Intelligence Agency Grant HM1582-05-C-0009.

References

1. S. M. Sherman. Tonic and burst firing: dual modes of thalamocortical relay. *TINS*, 24:122–126, 2001.
2. R. Krahe and F. Gabbiani. Burst firing in sensory systems. *Nature Reviews: Neuroscience*, 5:13–23, 2004.
3. M. Steriade and R. R. Llinas. The functional states of the thalamus and the associated neuronal interplay. *Physiol. Rev.*, 68:649–742, 1988.
4. A. M. Sillito and H. E. Jones. Corticothalamic interaction in the transfer of visual information. *Phil. Trans. R. Soc. Lond. B*, 357:1739–1752, 2002.
5. W. Guido and T. Weyand. Burst responses in thalamic relay cells of the awake behaving cat. *J. Neurophysiol.*, 74:1782–1786, 1995.
6. T. G. Weyand, M. Boudreaux, and W. Guido. Burst and tonic response modes in thalamic neurons during sleep and wakefulness. *J. Neurophysiol.*, 85:1107–1118, 2001.
7. N. A. Lesica and G. B. Stanley. Encoding of natural scene movies by tonic and burst spikes in the lateral geniculate nucleus. *J. Neurosci.*, 24:10731–10740, 2004.
8. E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, 24:1193–1216, 2001.
9. L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.*, 40:1489–1506, 2000.
10. L. Itti and C. Koch. Computational modelling of visual attention. *Nature Rev. Neurosci.*, 2:194–203, 2001.
11. S. Shipp. The brain circuitry of attention. *TICS*, 8:223–230, 2004.
12. H. B. Barlow. Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, 1961.
13. M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B*, 216:427–459, 1982.
14. J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.