

A Low-Power Imager and Compression Algorithms for a Brain-Machine Visual Prosthesis for the Blind

L. Turicchia, M. O'Halloran, D. P. Kumar, and R. Sarpeshkar*

Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA USA 02139-4307

ABSTRACT

We present a synchronous time-based dual-threshold imager that experimentally achieves 95.5 dB dynamic range, while consuming 1.79 nJ/pixel/frame, making it one of the most wide-dynamic-range energy-efficient imagers reported. The imager has 150x256 pixels, with a pixel pitch of 12.5 μ m x 12.5 μ m and a fill factor of 42.7%. The imager is intended for use in a brain-machine visual prosthesis for the blind where energy efficiency and power are of paramount importance. Such prostheses will also need to convey visual information to patients with relatively few electrodes and in a manner that minimizes electrode interactions, just as cochlear implants have accomplished for deaf subjects. To achieve these goals, we present a strategy that compresses visual information into the basis coefficients of a few image kernels that encode enough information to provide reasonably good image reconstruction with 60 electrodes. The strategy also uses time-multiplexed stimulation of electrodes to minimize channel interactions like the continuous interleaved sampling (CIS) strategy used in cochlear implants. Some of the image kernels that we employ are similar to the receptive fields observed in biology and may thus be natural to learn, just as cochlear-implant subjects have learned to reconstruct sound from a few filter basis coefficients.

Keywords: Neural prosthesis, visual prosthesis, neural stimulation, artificial vision, medical implant, brain-machine interface, low-power imager, wide-dynamic-range imager.

INTRODUCTION

According to the World Health Organization, in 2002 more than 161 million people were visually impaired, of whom 124 million had poor vision and 37 million were blind¹. Moreover, the economic impact of vision loss in the United States alone is estimated at \$68 billion a year². An artificial vision system will have a huge impact on society and preliminary results with visual-prosthesis devices currently under development have been encouraging.

Neural tissue can be stimulated by pulses of electrical current delivered using electrodes. Experiments have shown that visual sensation can be created by stimulating neural tissue along the visual pathway. Currently, the four main approaches to visual prostheses involve retinal implants^{3,4,5}, optic-nerve cuff electrodes⁶, cortical implants⁷, and most recently, dorsal lateral geniculate nucleus (LGN) implants⁸. There are pros and cons to stimulation in various sites. For example, a brain-machine visual prosthesis (cortical or LGN stimulation) is potentially useful in all common causes of blindness and has significantly more space available for implantation, while a retinal prosthesis is more severely space constrained and cannot be used in glaucoma-related causes of blindness. However, a retinal-implant surgery presents less risks than a brain surgery for a patient who is already blind. Our work is primarily targeted towards brain-machine visual prostheses, although it could be useful in other visual prostheses as well.

Regardless of the stimulation site, in order to develop a visual prosthesis we need to address four critical issues: (1) the overall system has to operate with low power to be small, portable, and to minimize tissue heating; low-power recording, processing, stimulation, and wireless circuits for brain-machine interfaces are discussed in⁹ and the references contained therein; (2) the imager used in such a system needs to have a wide dynamic range and needs to operate with very low power to be able to sense real-life images in an efficient fashion; (3) effective coding strategies need to be developed to deliver image information to the subject using relatively few electrodes; and (4) stimulation strategies need to be developed to minimize electrode interactions commonly found during stimulation of neural tissues.

* rahuls@mit.edu; phone 1 617 258-6599; fax 1 617 253-5210; <http://www.rle.mit.edu/avbs/>.

In this paper, we present a low-power imager and compression algorithms that are meant to address the issues outlined in the previous paragraph and that represent progress towards constructing a practical visual prosthesis. Section 2 describes the low-power imager and Section 3 describes the compression algorithms. Our 150x256 imager experimentally achieves 95.5 dB dynamic range, 37 dB peak signal-to-noise ratio (SNR), and consumes 1.79 nJ/pixel/frame, making it one of the most wide-dynamic-range energy-efficient imagers reported to date. In addition, this imager has a high fill factor of 42.7% and a pixel pitch of 12.5 μ m x 12.5 μ m. In Section 3, we present an image coding strategy for visual prostheses that uses time multiplexing and basis coefficients from image kernels to compress information from a visual scene onto 60 electrodes. Preliminary experiments with a visual prosthetic simulator¹⁰ show excellent face and text recognition performance with just 60 electrodes—a number already feasible with prostheses currently being tested¹¹. In Section 3, we also discuss how our strategy may be able to alleviate electrode interactions due to its use of a synchronous continuous interleaved sampling (CIS) strategy, which has proven to be very successful in cochlear implants¹². Asynchronous interleaved sampling (AIS) techniques for stimulation that minimize electrode interactions and that simultaneously lower stimulation power have been proposed for use in cochlear implants, especially for encoding music^{13,14}. Such AIS techniques are compatible with our strategy as well, but we only focus on the CIS version of our strategy in this paper.

Simulated reconstructions of the overall image after it has been processed with our compression-and-CIS strategy show fidelity that appears to be acceptable for landscape, face, and text recognition. The image kernels that we employ to extract basis coefficients are similar to those seen in receptive fields in biology and we can certainly use kernels that are identical to those measured in biology. Thus, while the overall success of our compression algorithm can only be judged after it has been tested on blind subjects, it does appear promising, and it is founded on ideas that have proven to be very successful in cochlear implants.

LOW-POWER, WIDE-DYNAMIC-RANGE IMAGER

The ideal visual prosthesis imager should have a wide dynamic range, minimize power consumption, and have a high fill factor. Conventional charge-coupled devices (CCDs) and CMOS active-pixel sensors (APs) have each found a niche within the field of digital imaging due to their complementary strengths in the areas of power consumption, image quality, and ease of system integration. However, the common shortfall of CCD and APS pixel arrays is their inability to faithfully capture images with wide intra-scene dynamic range. While the dynamic range of these sensors typically falls between 10³:1 (60 dB) and 10⁴:1 (80 dB), real-world scenes can exhibit illuminations spanning a range of 10⁵:1 (100 dB) or more¹⁵. When confronted with such a scene, these sensors either fail to capture significant information in the darkest areas of the image, saturate in the brightest areas, or both. The resulting information loss is not tolerable for a visual prosthesis in all light conditions. This has led the push to develop modified sensor topologies with improved dynamic range.

Besides having a wide dynamic range, a visual-prosthesis imager needs to have low power consumption. Low power consumption is important to ensure long-term viability of the visual prosthesis, as the device needs to operate with a small portable battery. Furthermore, a high fill factor is desirable to ensure that most of the silicon area is used for imaging rather than for overhead circuitry, thus enabling good spatial resolution and a small imager size.

Various CMOS APS imager topologies have been proposed to increase the dynamic range of each pixel. Among the potential solutions are time-based imagers. Time-based imagers encode photocurrent information in the amount of time it takes the sensed photocurrent to integrate a pixel capacitance to a globally defined threshold, rather than measuring voltage changes on this capacitance that result from a fixed time interval for photo-current integration. Time-based imagers offer the potential for high dynamic range in modern CMOS technologies since the dynamic range afforded by a time signal variable gets better as silicon technologies get faster, while the dynamic range afforded by a voltage signal variable gets worse as power-supply voltages get smaller. Unfortunately, many past time-based approaches have failed to achieve wide-dynamic-range capture in video-rate applications¹⁶. One time-based imager employs a time-varying threshold that begins each frame at a high level to maximize the charge quanta collected by brightly-illuminated pixels, and then decreases this threshold over the course of the frame to ensure that dimly-illuminated pixels also reach threshold within an allotted video frame time¹⁷. However, such imagers are not energy efficient and consume nearly 132nJ/pixel/frame. Our imager improves on prior time-based imagers, using four key innovations to solve several practical problems that have formerly limited the dynamic range and/or the energy efficiency of such imagers.

The four key innovations present in our time-based imager are as follows: First, a novel dual-threshold time-based current sensing algorithm is utilized that forces each single-slope integrating pixel to cross two threshold levels per frame (once just after reset and a second time after a near-optimal amount of photo-generated charge has been collected). This differential measurement technique eliminates offset fixed pattern noise (FPN) and pixel reset noise, and reduces comparator $1/f$ noise. Second, synchronous threshold detection is employed, yielding significant power savings compared with asynchronous approaches. Third, a method of optimizing the global dual-threshold waveform and associated pixel threshold-detection times is used. This method ensures that the quantization noise introduced by the algorithm remains negligible compared to the intrinsic pixel noise floor, while simultaneously minimizing the number of threshold detections employed, and thus the energy consumed. Fourth, a novel capacitively-coupled pixel topology is employed. This topology enables highly-linear responses to be achieved with this algorithm, while minimizing the common-mode input range of the pixel comparator, thus simplifying its design. Together, these innovations result in an energy-efficient imager with wide-dynamic-range and high fill factor. In this brief conference paper, we will only describe the key ideas and summarize the experimental performance of the imager. Further details will be reported in an upcoming journal publication that focuses on the circuit details behind the imager.

1.1 Pixel Topology

A high-level view of the proposed synchronous time-based dual-threshold pixel and its associated operating waveforms is shown in Fig. 1. The pixel topology shown in part (a) is composed of a synchronous auto-zeroing comparator, linear coupling capacitor (which is added to each pixel in order to linearize the pixel charge-to-voltage conversion¹⁸, photodiode, and pixel-to-memory communication block. It also includes a parasitic voltage-dependent capacitance between V_{pd} and GND . Part (b) of the figure illustrates a typical set of pixel waveforms over a single frame capture. The frame begins with an auto-zeroing reset phase, which establishes the initial condition $V_{pd}(t_{res,end}) = V_{ref}$ on the photodiode cathode and a low voltage level $V_{thresh}(t_{res,end})$ on the opposite terminal of C_{coup} . The first threshold crossing phase begins at $t_{res,end}$, during which V_{thresh} makes an initial upward step and then sweeps downward through a range of voltages over the interval from $t_{res,end}$ to $t_{off,end}$. The size of the initial step and the voltage sweep are chosen to ensure that all pixels will detect a threshold crossing during this phase, while the rate at which the comparator and pixel-to-memory communication block are triggered by S is as high as possible. This strategy allows the voltage sweep to occur in the minimum possible time while simultaneously minimizing the quantization noise incurred in measuring t_{off} , the first threshold crossing time. The second threshold crossing phase begins at $t_{off,end}$, where V_{thresh} steps up to its maximum level, coupling the maximum possible amount of charge into V_{pd} . For much of the remaining frame time V_{thresh} remains fixed at this peak voltage while the comparator and pixel-to-memory communication block are repeatedly triggered by S at a rate that exhibits a slowing trend over the course of the frame. In this particular example, the photocurrent is large enough to cause V_{pd} to cross V_{ref} during this at-top region of V_{thresh} . Based on the polling interval dictated by S in this region, the pixel-to-memory communication block will inform the pixel's associated memory element, via the column bus, that the second threshold crossing t_{thresh} occurred at some time between t_x and t_y , as shown in Fig. 1(b), causing a small quantization error. The dual-threshold waveform of V_{thresh} is optimized in order to minimize the quantization noise introduced by the algorithm and to ensure that it is negligible compared to the intrinsic pixel noise floor. The dual-threshold measurement algorithm that we have innovated is particularly beneficial because both t_{off} and t_{thresh} are measured using threshold crossings. Therefore, I_{photo} may be estimated¹⁸ without introducing significant offset errors caused by dynamic effects such as the reset switch charge injection and comparator latching mismatch that cannot be removed by auto-zeroing alone¹⁹.

If the photocurrent is too small, the second threshold crossing does not occur during the maximal threshold phase where V_{thresh} is flat and at a maximal value, and we are then unable to sense the photocurrent. To avoid this scenario, after a maximal amount of time has passed, V_{thresh} follows a steep downward descent while the comparator and communication circuits are again rapidly triggered by S , forcing the second threshold crossing to occur at some time during the descent. This scheme ensures that even low-light intensities will register a second threshold crossing within the frame period, thus guaranteeing a constant-frame-rate wide-dynamic-range imager.

Synchronous communication for pixel data read out was chosen because it lowers power consumption: In a synchronous communication scheme, the comparator needs to be turned on only during the readout period. In contrast, in an asynchronous communication scheme, as has been commonly used in prior time-based imagers, the comparator needs to be on at all times to register a threshold crossing as soon as it occurs, thus wasting power.

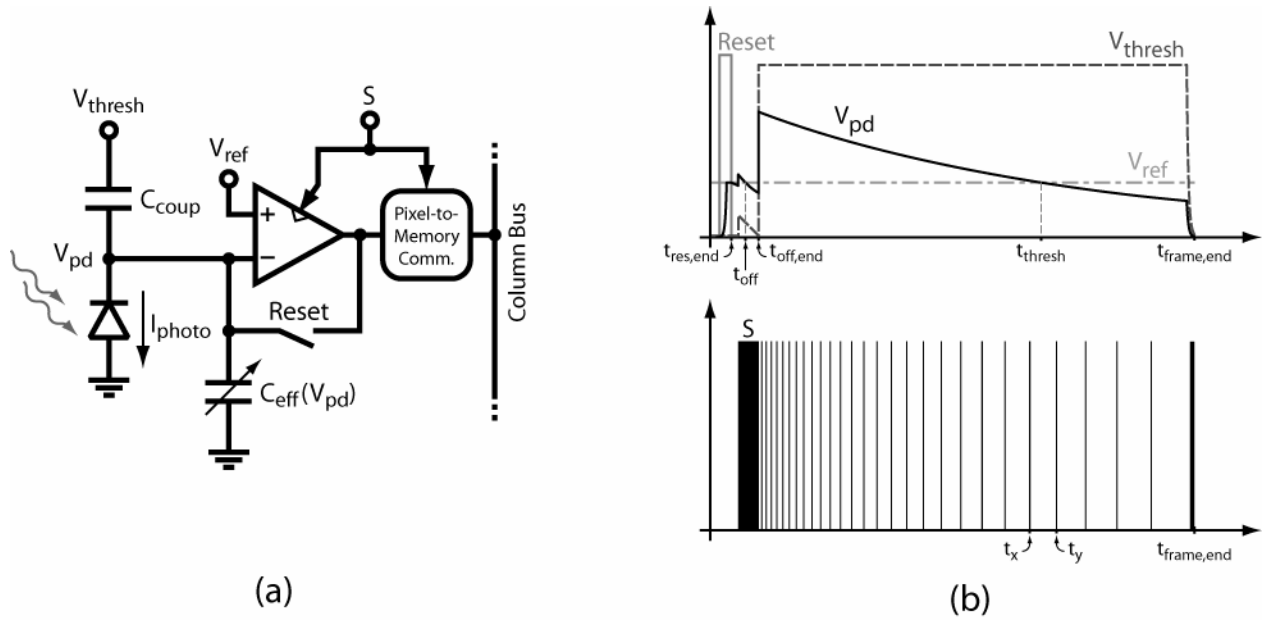


Fig. 1: Pixel topology (a) and representative operating waveform (b). Note the length of the reset and offset phases have been exaggerated to illustrate details within this time interval.

1.2 Experimental Results

A dual-threshold imager with 150 x 256 pixels was fabricated using a 0.18 μm CMOS 1.8V process. The die photo shown in Fig. 2 illustrates the major functional blocks of the imager. The overall die measures 5000 μm x 5000 μm including pads, enabling a pixel pitch of 12.5 μm x 12.5 μm with a 42.7% fill factor.

A plot of the measured array signal-to-noise ratio (SNR) versus illuminance, along with the response predicted by the noise theory described in¹⁸, is shown in Fig. 3. The plot demonstrates that the imager array experimentally achieves 95.5 dB dynamic range with 37 dB peak SNR. The individual pixels experimentally achieve 98.8 dB dynamic range and 44 dB peak SNR. The array performance lags slightly behind that of the individual pixels due to the additional noise power contributed to the array data by pixel-to-pixel mismatch effects. Our performance is quite competitive with good commercial imagers since it is nearly photon-count shot-noise limited as the data and theory of Fig. 3 reveal.

Besides dynamic range and SNR, the power consumed by an imaging array during frame capture is another important measure of its performance. To allow for direct comparisons of power consumption across various designs, the following metric was defined,

$$Energy_Efficiency = \frac{Total_Power_Consumed}{\#Pixels \times Frame_Rate}$$

which normalizes each array's power by the number of pixels it contains and its frame-rate. Based on experimental measurements, the prototype dual-threshold imager achieves an energy efficiency of 1.79 nJ/pixel/frame. This energy efficiency level is compared with other reported designs in the scatter plot shown in Fig. 4, which maps each design's achieved energy efficiency versus its reported array dynamic range, with more efficient implementations occurring towards the bottom-right corner of the figure. The plot demonstrates that the prototype dual-threshold imager is one of the most energy-efficient wide-dynamic-range imagers reported to date.

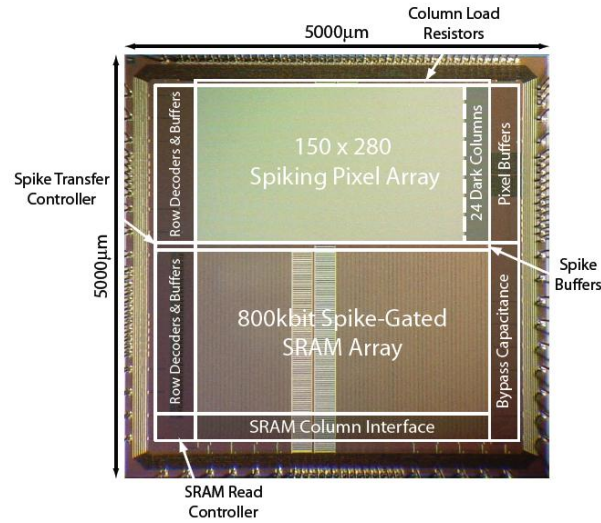


Fig. 2: Dual-threshold imager die photo

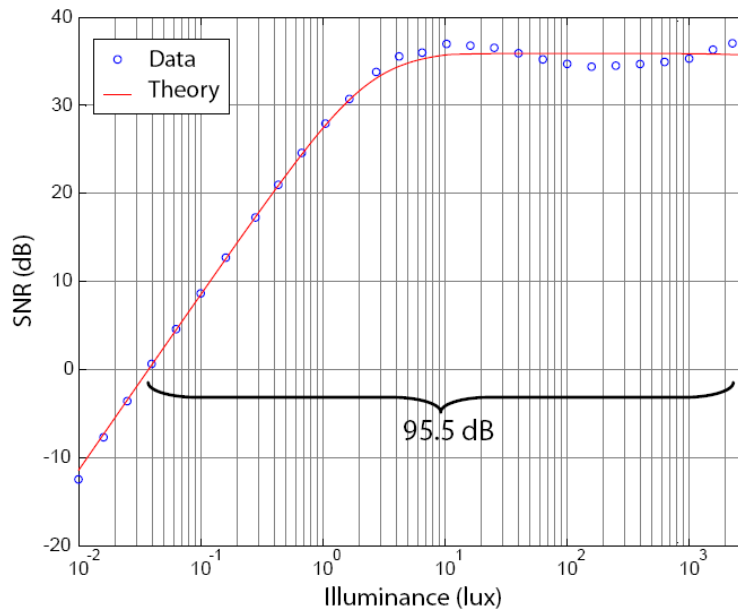


Fig. 3: The measured and theoretical array SNR characteristics. The array experimentally achieves 95.5dB dynamic range and 37 dB peak SNR.

Fig. 5(a) reveals an image obtained with our 95dB imager. Fig. 5(b) reveals an image obtained with the same circuits and optics but that uses a traditional low-dynamic-range 62dB image-sensing algorithm. We note that, in Fig. 5(a), the bright outdoor environment outside the window reveals the cars and buildings that are present while in Fig. 5(b), the outdoor environment is so bright that it has saturated the imager, and no details in the outdoor environment are discernible. In both cases, the indoor visual scene, which is significantly less bright, is correctly sensed. Thus, the wide-dynamic-range imager that we have built is capable of correctly sensing visual scenes with both dim indoor environments and bright outdoor environments simultaneously while imagers with lower dynamic range are not.

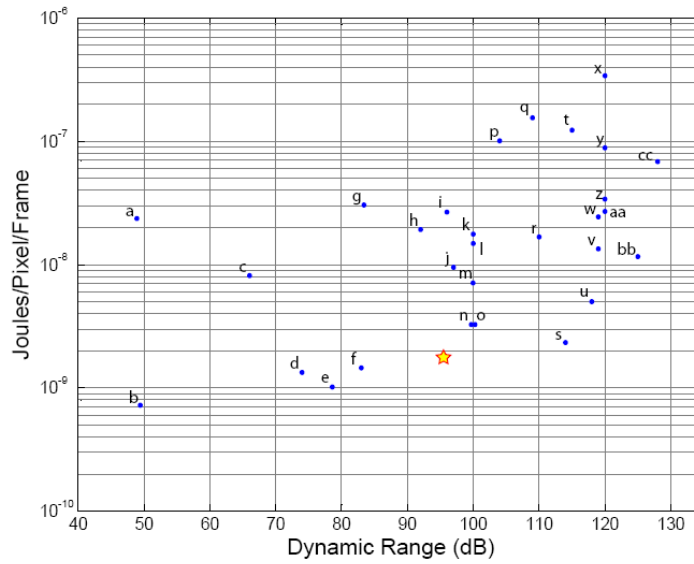


Fig. 4: Scatter plot of energy efficiency versus array dynamic range for previously reported imager designs. The star represents the performance achieved in the prototype dual-threshold imager. The lettered dots correspond to published results of other imagers. The references to each of these works can be found in¹⁸.



(a)



(b)

Fig 5. Part (a) reveals an image obtained with our 95dB imager. Part (b) reveals an image obtained with the same circuits and optics but that uses a traditional low-dynamic-range 62dB image-sensing algorithm.

IMAGE CODING AND COMPRESSION

In a visual prosthesis, the image obtained using a camera is usually directly coded into a relatively coarse spatial pattern of electrode stimulation that corresponds to the image, and with a monotonic mapping of image intensity to electrode charge. In this paper, we propose a strategy that codes this information in a compressed fashion such that the limited number of electrodes that are available are used wisely to convey information that is not just the mere pixel intensity but information that enables the brain to efficiently reconstruct the visual scene. The brain, which is the best unsupervised learning machine that we know, can then learn to interpret this information over time to reconstruct the visual scene. Is there evidence that the brain will be up to the task or able to interpret this information correctly? The success of cochlear implants gives us hope: It is actually quite astounding that cochlear implants work so well given that we have replaced 3500 spectral channels with 4-8 channels of spectral information. Cochlear implants work because the amazing brain is able to use the information present in the basis coefficients in time and in cochlear location effectively. The basis still provides complete information about the sound albeit in a more temporal rather than spectral fashion as the normal cochlea would do. Nevertheless, the brain is still able to learn to extract this information over a period of 3-6 months because it is there.

Fifty to hundred electrodes that convey mere image-intensity information do not provide enough information about the visual scene to reconstruct it. In fact, it has been shown that an image of at least 25×25 pixels is necessary for face recognition, mobility, and reading^{20,21,22}. Current visual prostheses have an electrode count that is significantly less than 625 and often in the 10-100 electrode range. Even if we could somehow increase electrode count, and somehow solve the attendant increase in power that comes with more electrodes, it is natural to wonder if a simple spatial mapping of image information to the electrodes makes the best use of the electrodes. Could a successful visual prosthesis be constructed with far fewer electrodes by preprocessing the image?

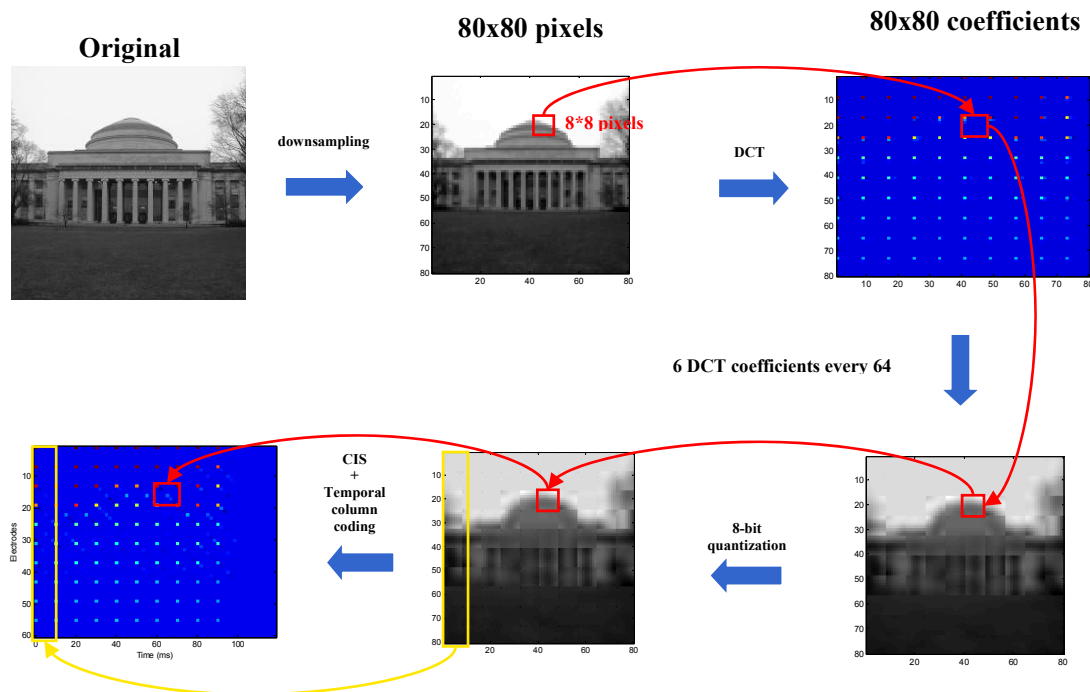


Fig. 6. Simplified diagram of the principal operations involved in one of our coding strategies.

Previous research has indicated that standard image-enhancement techniques, such as edge detection, are not useful in increasing the recognition of low-resolution images (25×25 pixels or below) in subjects with normal vision. Moreover, Dowling et al.²³ have shown that, with 50×50 images, Sobel edge-detection techniques and Canny edge-detection techniques perform worse than unprocessed images in picture-recognition tasks. These results undermine the efficacy of

standard image-enhancement techniques in visual-prosthesis systems. Interestingly, in cochlear-implant subjects, simple formant-extraction schemes that only provide formant information in a speech signal rather than the complete information available in a bank-of-basis-filters also perform poorly. Therefore, we have chosen an approach based on using image-kernel basis coefficients to compress and encode information in images. As we show later, our strategy leads to reasonably good reconstruction of the image. Sound-reconstruction techniques have been widely used in cochlear implants to test the efficacy of various processing strategies and have been found to be quite predictive of real performance in deaf subjects²⁴. Thus processing strategies that lead to acceptable image reconstruction may also suggest acceptable performance in blind subjects.

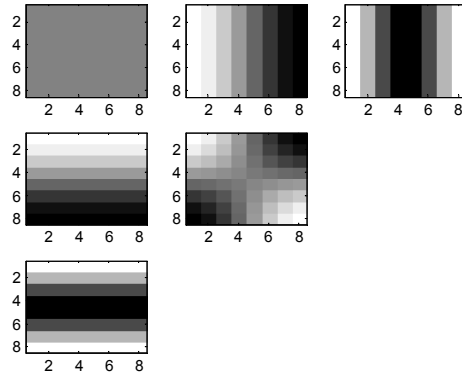


Fig.7. Squares corresponding to the low spatial-frequency templates into which the image is decomposed in a DCT compression strategy.

Fig. 6 shows a simplified diagram of the principal operations involved in one embodiment of our coding strategy. The video input from our imager is down-sampled or chopped to 80 x 80 pixels. The image is down-sampled in “mobility mode” (or when a wide-angle view is desired) and chopped in “reading mode” (when a zoomed view is preferred). The resulting image is split into blocks of 8 x 8 pixels, and for each block, the data undergoes a discrete cosine transform (DCT). The DCT transforms each pixel $i_{x,y}$ into a new coefficient $I_{u,v}$, using the following formula:

$$I_{u,v} = \sum_{x=0}^7 \sum_{y=0}^7 i_{x,y} \cos \left[\frac{\pi}{8} \left(x + \frac{1}{2} \right) u \right] \cos \left[\frac{\pi}{8} \left(y + \frac{1}{2} \right) v \right] \quad (1)$$

As a result, a new matrix of 80 x 80 coefficients is generated. Analogous strategies are commonly used for the compression of photographic images (e.g., JPEG). These methods usually also compress the coefficient information with appropriate quantization and Huffman encoding. In our strategy, we do not perform this step; instead, we use only the coefficients corresponding to the lowest six spatial frequency components, and we neglect the others, i.e., each 8 x 8 block is transformed to a linear combination of six DCT basis vectors or kernels. Since only 6 coefficients out of every 64 are used, our data-compression ratio is $64/6 \approx 11$. Fig. 7 shows the six basis kernels or vectors corresponding to the low spatial-frequency into which the image is decomposed. It is possible to reconstruct the image using only these components, as can be seen at the bottom right of Fig. 6. Next, these coefficients are quantized with an 8-bit quantizer and the corresponding reconstructed picture is visible at the bottom center of Fig. 6. These $6 \cdot 10 \cdot 10 = 600$ quantized coefficients encode information in the original image. This information is now sent to the stimulating electrodes of a visual prosthesis in a manner that minimizes electrode interactions.

We now discuss how compressed image information from an 80 x 80 pixel image may be conveyed to 60 electrodes in a fashion that minimizes electrode interactions: The image may be viewed as being composed of a square grid with 10 x 10 tiles, with each tile containing 8 x 8 pixels, and each tile being assigned to 6 electrodes, *when* it is chosen for stimulation. All 10 tiles in a given column in the grid are simultaneously chosen for stimulation and only one entire column is chosen for stimulation at a time such that no two columns can be active at once. Each tile uses a CIS strategy

as shown in Fig. 8 to stimulate 6 electrodes with the 6 DCT coefficients generated from an image analysis of that tile, one at a time in a non-overlapping sequence¹². Each electrode corresponding to that tile is therefore, dedicated, to a particular DCT coefficient, and only one DCT coefficient is stimulated at any one time to prevent electrode interactions amongst electrodes in that tile. Electrodes corresponding to other tiles of the same column in the grid; however, are separated by at least one tile distance, such that their simultaneous stimulation only causes weak interactions if any. Thus, at any one point in time, 10 electrodes in a given column, one in each tile, and all corresponding to the same DCT coefficient are stimulated. After all 6 x 10 electrodes corresponding to the 10 tiles of a given column have undergone stimulation, we repeat the procedure for the next column, then the next column, ...until all 10 columns have undergone stimulation, one at a time. Since only one column is active at a time, we reuse the same 60 electrodes for each round of columnar stimulation. Fig. 9 illustrates our overall scheme but, for clarity, only four tiles from the first two columns are drawn and all stimulation strengths are shown equal. In reality, the strengths of the DCT coefficients would be monotonically mapped to the electrode stimulation charge.

An electrode stimulation rate of 100 pulses per second (600 pulses per tile) implies a frame stimulation rate for 10 columns of $100/10 = 10\text{Hz}$. If we pause for 20ms after the last column has been stimulated in a given frame to permit the subject to identify the end of the image, the net frame rate is $1/(120\text{ms}) \approx 8\text{Hz}$. Thus, an 80 x 80 pixel image has been effectively coded into 60 electrodes by exploiting image compression and time multiplexing. Note that time multiplexing across columns saves us both electrodes and alleviates electrode interactions while time multiplexing within a tile only avoids interactions.

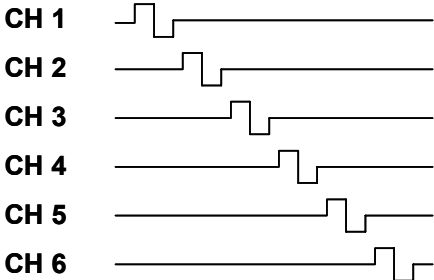


Fig. 8. A block of six coefficients is temporally coded using a CIS strategy. CIS allows only one electrode to fire at a time, avoiding channel interactions in that group.

Fig. 10 shows the original and a reconstructed picture using our strategy. Preliminary experiments with landscape, face, and text pictures (the most important kinds of images for the blind) show that our technique can create image reconstructions of reasonable fidelity for a visual prosthesis.

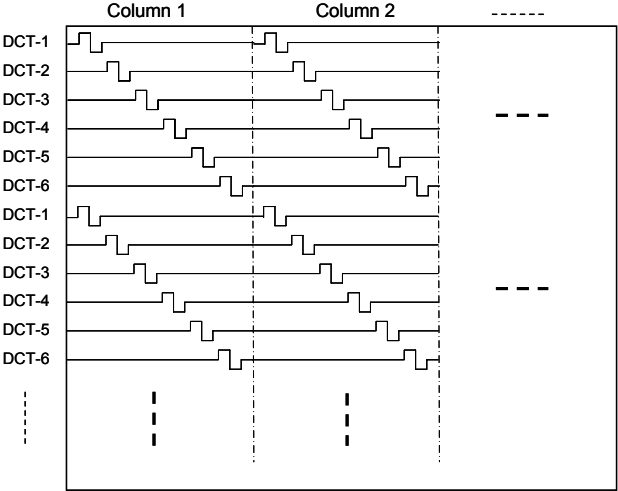


Fig. 9. Enlarged version of the bottom left of Fig. 6.

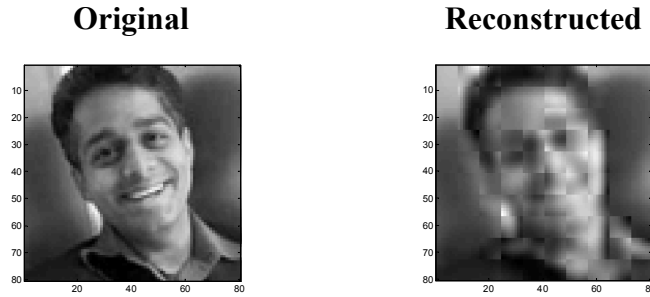


Fig. 10. The reconstructed image after compression shows that there is loss of fidelity as expected, but a low-resolution image suited for visual prostheses with 60 electrodes is feasible. Having more electrodes and more DCT coefficients can improve fidelity.

A blind subject with the visual prosthesis may perceive the sequential columnar presentation of image information as a sensation analogous to that experienced by a normal-vision subject as she walks past a door that is slightly ajar and sees, as she walks past the door, the entire visual scene of the room behind the door. A simple experiment performed on oneself shows that one can perceive all objects in a room in this fashion. The fast frame rate of 8Hz may suffice to ensure that the entire image is indeed reconstructed. The information related to which column is being encoded could be sent to the electrodes as well, allowing a better reconstruction of the picture by the subject.

In the strategy that we have thus far described, the image is coded column-by-column, but the coding can be done row-by-row, or by selecting a different subset of pixels at any discrete time instant. In fact, we can imagine a time-multiplexed strategy that performs no compression whatsoever but that just selects the subset of pixels that will be fed to the electrodes in a given discrete time instant. Thus, at any given instant, only a portion of the image is coded. The set of pixels that is simultaneously coded could be selected by translation of the previously selected pixels with a pre-determined movement (e.g., a cycloidal movement). A particular example of this technique is a strategy that codes the resized image of 80 x 80 pixels by selecting a different box of 20 x 3 pixels every 10 ms; the 20 x 3 pixels are located by moving this box every 10 ms. The sequence of these 60 pixels every 10 ms constitutes a temporal coding of the image, and the corresponding information is sent to 60 electrodes. Such a strategy does not require the brain to learn the compression kernels (the learning process is similar to that involved when cochlear-implant patients learn to understand sounds through a cochlear implant). Although, it is known that subjects wearing an experimental visual prosthesis usually scan the environment continuously by moving their head⁵, to our knowledge, an image time coding strategy that automatically scans the image has never been used. Interestingly, jumping spiders view their environment with a one dimensional retina that scans the world²⁵.

The spatial profile of simple-cell receptive fields in the macaque primary visual cortex can be extremely well modeled by two-dimensional Gabor functions²⁶. This finding suggests the use of Gabor kernels in our strategy when the electrodes of the visual prosthesis are used to directly stimulate the primary visual cortex. In this case, a set of kernels consisting of two-dimensional Gabor kernels with various scales, rotations, and positions is the best choice. The kernels are matched with the image of interest, which is then decomposed into a set of coefficients, similar to what we have previously shown with the DCT. The Gabor kernel is defined by

$$g(x, y, x_0, y_0, \lambda, \theta, \psi, \sigma, \gamma) = e^{-\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (2)$$

where

$$\begin{aligned} x' &= (x - x_0) \cos \theta + (y - y_0) \sin \theta \\ y' &= -(x - x_0) \sin \theta + (y - y_0) \cos \theta \end{aligned} \quad (3)$$

In this equation, λ represents the wavelength, θ represents the orientation, ψ is the phase offset, and γ is the spatial aspect ratio. The center of the kernel is set by x_0 and y_0 . Fig. 11 shows an example of a set of kernels with

$x_0 = 0, y_0 = 0, \psi = 0, \sigma = 1, \gamma = 2, \theta = (0, \frac{1}{4}\pi, \frac{1}{2}\pi, \frac{3}{4}\pi)$, and $\lambda = (8, 2, 1, 0.5)$. Although the efficacy of this strategy has yet to be verified experimentally, the similarity of these kernels with the data measured in the macaque primary visual cortex²⁶ strongly suggest that the Gabor kernel is well suited for coding of images in a visual prosthesis, especially for a cortical prosthesis.

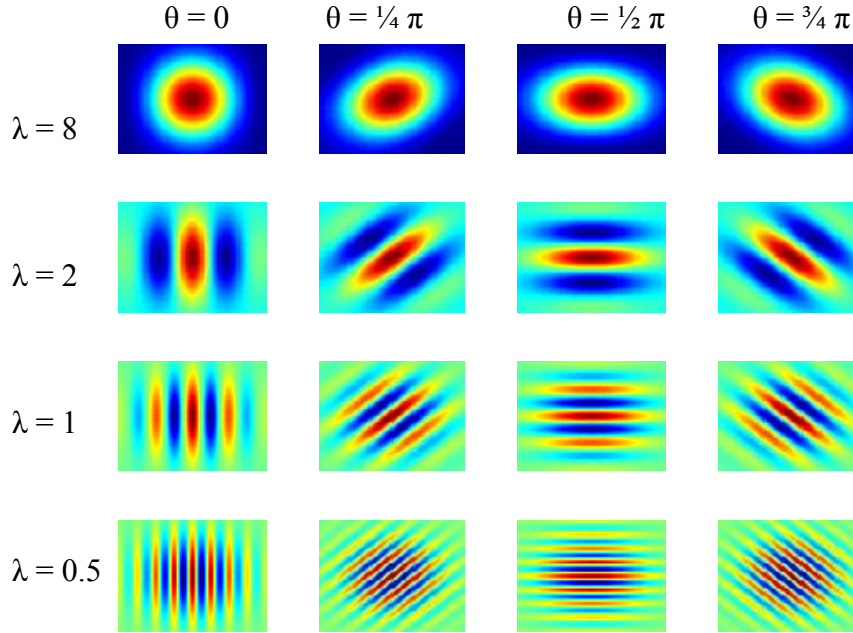


Fig. 11. Example of a set Gabor kernels with $x_0 = 0, y_0 = 0, \psi = 0, \sigma = 1, \gamma = 2, \theta = (0, \frac{1}{4}\pi, \frac{1}{2}\pi, \frac{3}{4}\pi)$, and $\lambda = (8, 2, 1, 0.5)$.

CONCLUSIONS

We discussed some steps towards accelerating the creation of a visual prosthesis for the blind. In particular, we discussed (1) the creation of a highly energy-efficient imager with wide dynamic range; (2) image coding, compression, and time-multiplexing strategies to minimize the number of electrodes needed for such a prosthesis; and (3) strategies to minimize electrode interactions. Our work is founded on the translation of several ideas from the mature field of cochlear implants that can be leveraged towards the development of a visual prosthesis for the blind. In fact, the utilization of time-multiplexed CIS strategies for minimizing electrode interactions and the utilization of filter-based kernel basis functions for compression have both resulted in significant improvements in the performance of cochlear implants. We suggest that they may hold similar promise for visual prostheses as well.

ACKNOWLEDGMENTS

The authors would like to thank Benjamin I. Rapoport for useful discussions. M. O'Halloran gratefully acknowledges support from the Catalyst Foundation. This work was supported in part by a grant from the McGovern Institute for Neuro Technology (MINT) program at MIT.

REFERENCES

- [1] World Health Organization, "WHO Fact Sheet N°282: Magnitude and causes of visual impairment," (2004).
- [2] Friedman, D. S., O'Colmain, B. J., Munoz, B., Tomany, S. C., McCarty, C., de Jong, P. T., Nemesure, B., Mitchell, P., and Kempen, J., "Prevalence of age-related macular degeneration in the United States," *Arch. Ophthalmol.*, 122, 564–572 (2004).
- [3] Wyatt, J., Rizzo, J., "Ocular implants for the blind," *IEEE Spectrum*, 33 (5), 47-53 (1996).
- [4] Liu, W., Sivaprakasam, M., Wang, G., Zhou, M., Weiland, J. D., and Humayun, M. S., "Development of an Intraocular Retinal Prosthesis to Benefit the Visually Impaired," in "Visual Prosthesis and Ophthalmic Devices: New Hope in Sight," edited by: Joyce Tombran-Tink, Joseph F. Rizzo, and Colin J. Barnstable, Humana Press Inc., Totowa, NJ (2007).
- [5] Weiland, J. D., and Humayun, M. S., "Visual Prosthesis," *Proceedings of the IEEE*, 96(7), 1076-1084 (2008).
- [6] Brelén, M. E., De Potter, P., Gersdorff, M., Cosnard, G., Veraart, C., Delbeke, J., "Intra-orbital implantation of the stimulating electrode for an optic nerve visual prosthesis : case report", *J Neurosurg.*, 104(4), 593-7 (2006).
- [7] Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K., and Vallabhanath, P., "Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex", *Brain*, 119(2), 507 (1996).
- [8] Pezaris J.S., Reid R.C., "Demonstration of artificial visual percepts generated through thalamic microstimulation," *Proc Natl Acad Sci USA*, 104(18), 7670-5 (2007).
- [9] Sarpeshkar, R., Wattanapanitch, W., Arfin, S. K., Rapoport, B. I., Mandal, S., Fee, M. S., Musallam, S., Andersen, R. A., "Low-Power Circuits for Brain-Machine Interfaces," *accepted for publication in IEEE Transactions on Biomedical Circuits and Systems*, (2008).
- [10] Hallum, L. E., Dagnelie, G., Suaning, G. J., and Lovell, N. H., "Simulating auditory and visual sensorineural prostheses: a comparative review," *J. Neural Eng.* 4, S58–S71, (2007).
- [11] Second Sight Medical Products Inc, "Second Sight Completes U.S. Phase I Enrollment and Commences European Clinical Trial for the Argus II Retinal Implant," press release, February 15, 2008. <http://www.2-sight.com/press-release2-15-final.html>
- [12] Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M., "Better speech recognition with cochlear implants," *Nature* 352, 236 - 238 (1991).
- [13] Sit, J., Simonson, A. M., Oxenham, A. J., Faltys, M. A., and Sarpeshkar, R., "A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information", *IEEE Transactions on Biomedical Engineering*, 54, 138-149 (2007).
- [14] Sit, J., and Sarpeshkar, R., "A Cochlear-Implant Processor for Encoding Music and Lowering Stimulation Power," *IEEE Pervasive Computing—special issue on implantable electronics*, 1(7), 40-48 (2008).
- [15] Debevec, P. E., and Malik, J., "Recovering high dynamic range radiance maps from photographs," in *Proc. Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, Los Angeles, CA, USA, 369-378 (1997).
- [16] Harris, J.G., "The changing roles of analog and digital signal processing in CMOS image sensors," *Acoustics, Speech, and Signal Processing*, 2002. *Proc. IEEE International Conference ICASSP '02* 4, 3976-3979 (2002).
- [17] Guo, X., Qi, X., and Harris, J., "A time-to-first-spike CMOS image sensor," *IEEE Sensors J.*, 7(8), 1165-1175 (2007).
- [18] O'Halloran, M., "A Wide-Dynamic-Range Time-Based CMOS Imager," Ph.D Thesis, Electrical Engineering, MIT, Cambridge, MA, USA (2008).
- [19] Qi, X., Guo, X., and Harris, J., "A time-to-first spike CMOS imager," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS'04)*, vol. 4, Vancouver, BC, 824-827 (2004).
- [20] Cha, K., Horch, K., and Normann, R.A., "Simulation of a phosphene-based visual field: visual acuity in a pixelized vision system," *Ann Biomed Eng.* 20(4), 439-49 (1992).
- [21] Cha, K., Horch, K. W., Normann, R. A., and Boman, D. K., "Reading speed with a pixelized vision system," *J Opt Soc Am A.*, 9(5), 673-7 (1992).
- [22] Cha, K., Horch, K. W., and Normann, R. A., "Mobility performance with a pixelized vision system," *Vision Res.* 32(7), 1367-72 (1992).
- [23] Dowling, J. A., and Maeder, A. J., "Mobility enhancement and assessment for a visual prosthesis." In Amini, Amir A. and Manduca, Armando, Eds. *Proceedings SPIE Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*, 780-791, San Diego (2004).

- [24] Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X., "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* 110, 1150 (2001).
- [25] Land, M. F. "Movements of the Retinae of Jumping Spiders (Salticidae: Dendryphantinae) in Response to Visual Stimuli" *J Exp Biol* 51:471-493 (1969).
- [26] Ringach, D. L., "Spatial Structure and Symmetry of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex," *J Neurophysiol* 88, 455-463 (2002).