Augmentations to enhance perception in prosthetic vision (also known as bionic eyes) have the potential to improve functional outcomes significantly for implantees. In current (and near-term) implantable electrode arrays resolution and dynamic range are highly constrained in comparison to images from modern cameras that can be head mounted. In this paper, we propose a novel, generally applicable adaptive contrast augmentation framework for prosthetic vision that addresses the specific perceptual needs of low resolution and low dynamic range displays. The scheme accepts an externally defined pixel-wise weighting of importance describing features of the image to enhance in the output dynamic range. Our approach explicitly incorporates the logarithmic scaling of enhancement required in human visual perception to ensure perceivability of all contrast augmentations. It requires no pre-existing contrast, and thus extends previous work in local contrast enhancement to a formulation for general image augmentation. We demonstrate the generality of our augmentation scheme for scene structure and looming object enhancement using simulated prosthetic vision.

1 Introduction

The ability to perceive and discern relevant structural features is critical to safe and efficient navigation, and object interaction. However, these tasks are made challenging when visual function is lost or severely compromised. In particular, degraded visual acuity and/or lost contrast perception increases the risk of missing important visual features cueing the presence of obstacles, trip hazards, drop-offs, and other objects of interest in the environment.

Numerous mobility aids have been proposed to alleviate lost visual function. Most common among these are traditional devices such as the long cane, and guide dogs. More recently, however, new technologies have emerged that aim to restore a sense of visual perception through alternative modes of stimulation. In this paper we focus on prosthetic vision.

Visual prostheses aim to elicit visual perceptions by way of electrical stimulation of healthy neurons; most commonly in the retina [12, 29]. While blindness causing diseases such as Retinitis Pigmentosa and Age-related Macular Degeneration progressively destroy photoreceptor cells, other cell layers of the retina remain largely intact, allowing electrical stimulation to occur. The percept elicited by this process is known as a phosphene: described as a bright ‘star-like’ spot of light [2]. The brightness and size of phosphenes may be manipulated using different stimulation parameters to produce a patterned image. Several retinal prostheses are currently undergoing clinical testing, including one device now with FDA approval to sell [12].

The dominant approach for capturing scene data is via an externally worn camera. Current implants utilizing external camera systems, however, only provide up to 60 electrodes, and up to around 10 brightness levels in the perceived output dynamic range. Thus, the task is to transform high resolution images to an encoding of the scene suitable for stimulation across a highly limited bandwidth. In most cases, this is achieved via a simple down-sampling of the captured image. However, this limits the use of the device to relatively simple, high-contrast environments.

The availability of high resolution digital images gives rise to the possibility of applying image processing and computer vision to improve perception in the output display. Amongst the most general of these include methods for performing contrast enhancement [4, 19, 25], in which existing intensity gradients are increased to enhance the perception of shape and form. Such approaches have been shown to improve image quality for visually impaired viewers [24, 25], but do not address the need for emphasis of low-contrast objects and features in a very low resolution and reduced dynamic range display. Moreover, not all contrast in the scene is relevant to tasks such as mobility or object interaction, where the key information required relates to the geometric structure of the environment. Thus, a key component of any enhancement under constrained viewing conditions is the selective enhancement of features of relevance to the task.

In the context of low vision aids, numerous algorithms have been proposed to detect key features in the scene, and/or segment areas of importance with respect to particular tasks (e.g., the ground plane [21], step detection [10], objects of interest [6, 11, 20, 6], etc.). While perceptual considerations of augmented displays have been extensively considered for augmented and mixed reality systems [15, 9], less attention has been given to the question of how to augment contrast to ensure the perception of important features in a greatly diminished output display, as is presently the case for prosthetic vision. In most cases, augmentations are crafted for specific tasks, and with particular display constraints in mind. In so doing, there is an inevitable trade-off between the augmentation performed and its generality to other activities of daily living. While a device may provide multiple user-selected modes, it is preferable to have a single consistent augmentation scheme that allows as much of the naturally occurring scene to pass through as possible.

In this paper we propose an adaptive augmentation framework for enhanced perception with low resolution/dynamic-range visual prosthetics. We assume that image data is first obtained via a high resolution body-worn sensor, along with a precomputed importance map providing a weighting of importance (either binary, or across a continuous range) for each pixel in the input image. The framework provides two mutually exclusive augmentations:

1. local contrast enhancement of at least a just noticeable difference to the background in regions of high importance

2. local contrast attenuation in regions of low importance to ensure key features take prominence in the reduced dynamic range.
The distinction between high and low importance is determined from a user specified threshold. These augmentations are formulated as local adaptive contrast adjustments in a logarithmic image processing framework, and in accordance with Weber’s law to ensure perceivable contrast change when the importance of the pixel is above a given threshold. We demonstrate the effectiveness and generality of our augmentation framework using importance maps derived from: (1) a recently proposed depth-saliency algorithm to enhance points of structural interest in the scene, and, (2) a time-to-contact estimation method to facilitate the enhancement of objects posing an imminent threat of collision.

The paper is structured as follows. Section 2 briefly overviews relevant literature. Section 3 outlines the proposed augmentation framework. Section 4 validates our augmentation using synthetic images. Section 5 evaluates and demonstrates the effectiveness of our augmentation for prosthetic vision using importance maps obtained from a depth-saliency algorithm, and time-to-contact estimator. Section 6 presents our conclusions.

2 Previous work
The perceptual requirements of augmented displays has been a focus of attention of augmented reality (AR) research for some time [15, 9]. In most cases, these considerations are with respect to the natural fusion between real and virtual content such as for occluded object visualisation [28] and/or improving relative depth judgement of inserted virtual objects [17]. Julier et al. [14], for example, propose a region-based filter for mobile augmented reality systems that account for user location (via tracking) and a priori defined user objectives. The filter adjusts object transparency in accordance with its inferred importance. Sandor et al. [28] use visual saliency to determine the relative contributions of features to display when fusing two displays. Hue, luminosity and edge overlays are used to enhance relevant features in the scene.

We focus specifically on the augmentation of contrast for a reduced resolution and dynamic range display. Adaptive contrast enhancement is well known, with Lee [19] proposing a model that adds contrast above the local mean according to a contrast scale parameter. It is known that for human viewers, the noticeability of enhancements scales logarithmically [23, 27] due to adaptation in perception. However, when dynamic range on the output display is constrained, augmentations need to be efficient. Thus, we replace the user defined scale parameter, and introduce just noticeable difference and exponential scaling to image augmentation to allow us to fix all parameters based on a model of the user’s ability to perceive contrast on the display. While previous work such as [13] have incorporated the requirement for just noticeable difference into contrast enhancement with respect to local image noise and image detail, here we further base our enhancement on an externally defined concept of importance for an augmentation that describes when we need to ensure at least a just noticeable difference, and by how much more to ensure it’s relative prominence in the final display. Moreover, our approach does not assume the pre-existence of contrast in regions of importance, and thus extends the concept to a generally applicable formulation for image augmentation.

3 Approach
3.1 Problem statement and formulation
Given an input image, \(I\), containing \(n\) image points and \(d\) brightness levels, and an output display \(I'\) with \(n\) points and \(d'\) brightness levels where \(d' < d\), the objective is to assign values to \(I'\) such that features of highest priority in \(I\) are preserved and/or enhanced in the output display. We assume a precomputed map \(S\), defining a priority weighting in the range \([0, 1]\) for each input pixel in \(I\).

We formulate the augmentation via the combined enhancement of local contrast in the vicinity of features of interest, and dampening of contrast in regions of reduced importance. The intuition of our approach is to give prominence and contrast to regions of the image deemed to be of most relevance to tasks such as safe mobility and general object interaction (e.g., highlight an object of interest on top of a large smooth surface, a trip hazard on the ground). We discuss example priority maps in Section 5.

We define our augmentation using the adaptive contrast enhancement equation originally proposed by Lee [19]:

\[
I_p' = \mu_p + (1 + \beta)(I_p - \mu_p),
\]

where \(\mu_p\) is the local mean intensity of \(I\) in the neighbourhood of \(p\), \(I_p\) is the intensity at \(p\), and \(I_p'\) is the output intensity value. The manually set scale factor \(\beta\) determines the aggressiveness of contrast adjustments. Following Deng et al. [5], we formulate our augmentation using the logarithmic image processing (LIP) form of (1), which has been proven to provide a better characterisation of saturation and other properties of the human vision system [23, 27]. Thus, Equation 1 becomes:

\[
\log(I_p') = \log(\mu_p) + (1 + \beta)(\log(I_p') - \log(\mu_p)),
\]

where \(I'\) and \(I\) are each defined as:

\[
I = 1 - \frac{I}{I_{\max}},
\]

and \(I_{\max}\) is the maximum allowable value in the dynamic range of \(I\). Henceforth we assume \(I\) and \(I'\) are normalised to the range \([0, 1]\).

The logarithmic mean intensity, \(\log(\mu_p)\) is defined as the mean of \(\log(I_p)\). Once computed, the result of Equation 2 can be transformed back to a linear scale via:

\[
I' = 1 - \exp(\log(I')).
\]

3.2 Contrast augmentation
We define our proposed augmentation by replacing the second term in Equation 2 with our proposed contrast adjustment function:

\[
\log(I_p') = \log(\mu_p) + \text{sign}(\Delta_t)\Delta_p
\]

where \(\Delta_p\) defines the extent of desired contrast change for two distinct cases:

\[
\Delta_p = \begin{cases} 
\max(|\log(1 - S_p')|, |\Delta t|) & \text{if } S_p > s_t \\
\max(|\Delta t| - |\Delta s|, 0) & \text{otherwise}
\end{cases}
\]

and:

\[
\Delta_t = \log(I_p') - \log(\mu_p),
\]

\[
\Delta s = \log(1 - s_t^2) - \log(1 - S_p').
\]

For \(S_p > s_t\) (i.e., the weighted importance of \(I_p\) is above threshold), the objective is to increase contrast, or if sufficient contrast exists, retain the existing contrast. Conversely, for \(S_p < s_t\), the objective is to reduce contrast in proportion to the relative difference between \(S_p\) and \(s_t\). The term, \(\text{sign}(\Delta t)\), ensures contrast adjustments follow the existing intensity gradient direction. Note that the manually set contrast gain, \(\beta\) from Equation 2 has now been replaced with a parameter \(\gamma\) that, as we will show, we fix on a model of human contrast perception that is particular to the display, and \(s_t\), which comes from the importance map.

3.2.1 Ensuring perceivable contrast
For the case that \(S_p > s_t\), the desired increase in augmented contrast is achieved by a power law which follows the model of perceived contrast and physical contrast in human perception, above low luminance [26], hence the parameter is exponential with respect to the mean intensity of the region in the discretised output dynamic.
range. We apply the Weber-Fechner Law for determining the required change in input stimulus required to achieve a just noticeable difference in the perceptual response. The law encapsulates the psychophysically observed property that the threshold for perceiving a just noticeable difference, $\delta'$, increases proportionally with the stimulus magnitude, such that [8]

$$\delta' = K_w (V_o + V),$$

(9)

where $V$ is the instantaneous input stimulus, and $V_o$ is a threshold input stimulus for a lowest perceivable response, $K_w$ is the empirically determined Weber fraction, which determines the sensitivity to response changes in a specific human viewer.

To ensure perceivability of contrast adjustments from the mean intensity of the local neighbourhood, we define $\delta'$ as:

$$\delta' = K_w \left( \mu_p + d_p' \right),$$

(10)

where $d_p$ is the minimum perceivable intensity value in the output dynamic range when the importance is greater than $s_t$. Given the just noticeable threshold, $\delta'$, we then solve for $\gamma$ such that:

$$s_t^p = \delta',$$

(11)

from which we obtain:

$$\gamma = \frac{\log(\delta')}{\log(s_t)}.$$  

(12)

Thus, for a given local mean intensity $\mu_p$ and weighted importance, $s_t$, the resulting contrast increment objective, $s_t^p$ in Equation 6 will be greater than $\delta'$, for all $s_p > s_t$.

### 4 Validation

To validate the properties of the proposed augmentation scheme, a hand-labelled importance map was generated. Figure 1 shows an importance map consisting of two concentric rings, the outer of value $s_t = 1$, and the inner set to $s_t = 0.5$. All other importance weights were set to zero. The proposed augmentation algorithm was applied to each of the input images shown in Figures 1(a-d) using the synthetic importance map as input. For these experiments, a $15 \times 15$ local averaging window was used, $K_w = 0.08$ and the output dynamic range was set to 8 brightness levels.

The resulting augmentations for $s_t = 0.1$ and $s_t = 0.5$ are given in the middle and last columns of Figure 1. Note that the input images for Figures 1(a-c) all contain no contrast. The resulting augmentations, however, produce noticeable contrast changes in accordance with the importance map, and the set threshold. Note that $s_t = 0.5$ augmentations allow no contrast augmentation to result from the $s_t = 0.5$ importance ring, confirming the desired behaviour that sub-threshold importance values do not contribute to an increased contrast. Figures 1(a) and (b) also confirm the symmetry of contrast enhancements when the image is uniformly black or white. Corresponding plots of contrast change across the middle row of each augmented image are given in Figures 2(a-c), confirming the visual contrast changes visualised in Figures 1(a-c).

Figure 1(d) shows the resulting augmentation in the presence of a uniformly grey striped pattern. As above, $s_t = 0.1$ allows large contrast enhancements to occur at all points where non-zero importance weights exist. Given the low threshold, augmentations generally result in an increase on existing contrast. Close inspection also shows less augmentation resulting from the $S_p = 0.5$ ring. As before, for $s_t = 0.5$, only the outer ring produces a noticeable contrast change in the image. Figure 2(d) again confirms these results in plots of intensity across the middle row of both augmented images.

![Figure 1: Examples of contrast augmentation achieved for the importance map shown above. Left column shows input images; middle and right columns show the output contrast augmentation for $s_t = 0.1$ and $s_t = 0.5$. Local intensity averages were computed within a $15 \times 15$ window, $K_w = 0.08$ and $\gamma = 0.8$.](image)

**Table 1:** EMEE results for $16 \times 16$ block regions above and below the importance threshold $s_t$.

<table>
<thead>
<tr>
<th>scene</th>
<th>orig</th>
<th>hist eq</th>
<th>local adaptive</th>
<th>proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$&gt; s_t$</td>
<td>$\leq s_t$</td>
<td>$&gt; s_t$</td>
<td>$\leq s_t$</td>
</tr>
<tr>
<td>tabletop</td>
<td>0.62</td>
<td>0.33</td>
<td>0.83</td>
<td>0.56</td>
</tr>
<tr>
<td>office</td>
<td>0.86</td>
<td>0.38</td>
<td>1.41</td>
<td>0.61</td>
</tr>
<tr>
<td>dropoff</td>
<td>1.02</td>
<td>0.34</td>
<td>0.89</td>
<td>0.27</td>
</tr>
</tbody>
</table>

### 5 Evaluation with Simulated Prosthetic Vision

We focus on the application of our proposed augmentation framework to prosthetic vision, using simulated prosthetic vision. Simulated prosthetic vision (SPV) aims to approximate the visual perception experienced by implanted patients, allowing for testing of vision processing strategies to occur with normal sighted participants. Simulation models commonly represent phosphenes as 2D Gaussians, with brightness and extent determined by the sampled intensity value [3]. While not a complete representation of the true perceptual experience, it provides a reasonable approximation to reported percepts in clinical trials [30, 22]. SPV provides a useful platform for evaluating and developing plausible image augmentation strategies without the need for patients, and without bias to a particular device design.

#### 5.1 Scene structure enhancement

Feng and McCarthy [7] propose a real-time depth-saliency algorithm designed to detect regions of structural change in a depth image (obtained either from stereo algorithms or consumer RGB-D cameras). The algorithm applies a multi-scale sliding window,
comparing histograms of depth gradient direction with surrounding regions to determine the uniqueness of the local gradient. The result is a so-called perturbance map, indicating the extent of structural change relative to local surrounding regions. Figures 3-5(b) shows example output for each corresponding image.

The proposed augmentation scheme was run over captured RGB-D image sequences, with augmentations performed on single channel intensity images and importance maps determined directly from the depth-saliency algorithm output. Sample images and corresponding importance maps are shown in Figures 3-5(a) and (b). The augmentation was applied with a $15 \times 15$ local mean intensity window and $K_w = 0.08$. For comparison, we also provide results for histogram equalisation and the logarithmic image processing (LIP) local adaptive method of [5] defined in Equation 2.

Figures 3-5(c-e) show the resulting images for all methods under comparison at full resolution, with 8 brightness levels. Figures 3-5(e-h) show corresponding SPV renderings for each method over a range of resolutions (the number of phosphenes) and dynamic ranges (the number of perceivable brightness levels).

Results in Figure 3 show a clearer visualisation of the lower contrast white objects on the light gray table-top surface using the proposed augmentation scheme. In particular, the two white cups which are given high importance via the perturbance map, are detectable in all but the top left SPV rendered output of the proposed augmentation method. At the most diminished display settings ($16 \times 12$ with 2 brightness levels), the cups and stapler remain perceivable. While both histogram equalisation and local adaptive methods retain some detail, the distinction between the table and objects appears comparatively worse at all corresponding display settings.

Figures 4 and 5 show examples of the augmentation in the navigation context. In both cases, histogram equalisation does a reasonable job of enhancing structural change where there is pre-existing contrast given sufficient dynamic range. However, the proposed augmentation provides a clearer delineation of key structural features such as the wall-floor boundaries in Figure 4, and the precise drop-off point of the staircase in Figure 5; at low dynamic range in particular. Note that in the latter case the most prominent contrast change marking the staircase drop-off in the original and histogram equalised images occurs three steps down, thus missing the crucial boundary between floor and staircase.

5.1.1 Quantitative evaluation

To provide an objective measure of contrast enhancements, we apply the measure of enhancement by entropy (EMEE) [1]

$$\text{EMEE}(e) = -\frac{1}{k_1k_2} \sum_{m=1}^{k_1} \sum_{l=1}^{k_2} \alpha(X)\alpha \log(X),$$  \hspace{1cm} (13)

where $X = \frac{I_m}{I_{max}}$. We choose this metric on the basis of it’s accepted and proven effectiveness as a human vision based metric of enhancement. For our comparison we set $\alpha = 1$ and $k_1, k_2 = 16$. Table 1 reports EMEE results for each method over full resolution (8 level) images shown in Figures 3-5. To assess the effectiveness of the proposed augmentation to increase contrast in important regions, and attenuate contrast in less important regions, we show

\footnote{a small offset is added to intensity values to avoid divide-by-zero error.}
Figure 4: Evaluation on office scene showing (a) the original image, and at full resolution: (b) the input importance map, (c) histogram equalisation result, (d) LIP local adaptive result ($\beta = 1$), (e) the proposed augmentation ($s_t = 0.2$). (f-h) show simulated prosthetic vision renderings for each of the above methods respectively. See Table 1 for quantitative comparison.

EMEE results computed for image blocks with mean importance above $s_t$, and for blocks below this threshold only.

EMEE results generally reflect the qualitative assessment above. The proposed method achieves the highest EMEE score for above importance threshold regions in two of the three images. The local adaptive method achieves a slightly higher EMEE score for the dropoff scene, however, as noted earlier, Figure 5 shows the proposed augmentation providing a more precise demarcation of the actual drop-off location (particularly as dynamic range drops). For regions below the importance threshold, Table 1 shows a substantial reduction in EMEE scores for the proposed method, confirming the attenuation of contrast away from important regions. We note, however, that such metrics provide guidance only. User studies in the context of real tasks are required to assess the true effectiveness of the augmentation scheme.

Figure 5: Evaluation on dropoff scene showing (a) the original image, and at full resolution: (b) the input importance map, (c) histogram equalisation result, (d) LIP local adaptive result ($\beta = 1$), (e) the proposed augmentation ($s_t = 0.2$). (f-h) show simulated prosthetic vision renderings for each of the above methods respectively. See Table 1 for quantitative comparison.

5.2 Looming object enhancement

To further demonstrate the generality of the proposed augmentation, we applied the method to an image sequence with corresponding time-to-contact maps providing for each pixel, a relative estimate of time before impact (up to scale) with surfaces in the scene. It is well known that time-to-contact may be inferred from diverging optical flow patterns [16], and plays an important role in visuomotor control in a number of biological vision systems, including human vision [18].

Figure 6 shows representative example results from the sequence using the proposed scheme, and histogram equalisation. The augmentation is clearly evident across the body of the walker as he approaches the camera. Note that the contrast augmentation also increases as the walker gets closer, providing an interpretable cue for looming motion. Significantly less contrast is evident in histogram equalised images.
6 Conclusion
We have proposed a novel locally adaptive contrast augmentation scheme suitable for enhancing key features of importance with a low resolution, low dynamic range display such as a visual prosthesis. We have outlined a model for augmentation that utilizes an externally defined importance map to explicitly ensure features above a given weight of importance are always perceivable from the mean intensity of the background, and increases contrast in accordance with the weighted importance of each feature. Our model is consistent with known properties of the human vision system, in particular, the logarithmic scaling of enhancement required to achieve perceivable difference. Our results validate this model, and demonstrate the general applicability of the scheme to prosthetic vision enhancement, and low resolution/low dynamic range displays generally. Future work will evaluate the performance of the augmentation scheme in human trials using SPV and clinical evaluation with implanted patients.

Acknowledgements
NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program. This research was also supported in part by ARC through its Special Research Initiative (SRI) in Bionic Vision Science and Technology grant to Bionic Vision Australia (BVA).

References


