Bioinspired Robotic Vision with Online Learning Capability and Rotation-Invariant Properties

Dan Berco* and Diing Shenp Ang*

Reliable image perception is critical for living organisms. Biologic sensory organs and nervous systems evolved interdependently to allow apprehension of visual information regardless of spatial orientation. By contrast, convolutional neural networks usually have limited tolerance to rotational transformations. There are software-based approaches used to address this issue, such as artificial rotation of training data or preliminary image processing. However, these workarounds require a large computational effort and are mostly done offline. This work presents a bioinspired, robotic vision system with inherent rotation-invariant properties that may be taught either offline or in real time by feeding back error indications. It is successfully trained to counter the move of a human player in a game of Paper Scissors Stone. The architecture and operation principles are first discussed alongside the experimental setup. This is followed by performance analysis of pattern recognition under misaligned and rotated conditions. Finally, the process of online, supervised learning is demonstrated and analyzed.

1. Introduction

Convolutional neural networks (CNNs) are a subclass of artificial neural networks (ANNs), usually implemented as software algorithms constructed from multilayered perceptrons.[1,2] CNNs are widely used to analyze visual data by classification while tagging elements within an image. The large connectivity inherent to these networks allows for a high degree of fitting. CNNs are also known to have shift-invariant properties.[3,4] This means that the algorithm may be trained to classify elements regardless of their spatial translations in the input image. However, rotational invariance has proven to be a nontrivial challenge for CNNs.

In this sense, artificial vision is at a disadvantage compared with vision in living creatures whose survival depends on reliable, real-time image perception, regardless of facial orientation or body posture, especially considering the interplay between predator and prey. Some workarounds have been suggested.[5–9] However, they rely on large computational effort and offline preprocessing. An innovative approach to address this issue may be found in the development of bioinspired vision systems.

Considering the eye’s cellular structure, the mammalian retina is made of layered light-sensitive rods and cones that convert photonic stimulation into neural activity. Visual data are conveyed by synaptic junctions of bipolar and horizontal cells via neurotransmitter signaling.[10–12] An off-center bipolar cell displays a hyperpolarization-like response under illumination that decays into a depolarized state in darkness. An on-center bipolar produces a depolarized output under illumination and hyperpolarizes in darkness. In this manner, horizontal-cell neurotransmitter releases under “center-surround” configurations allow for better accuracy and resolution during real-time image acquisition. This functionality was emulated, through spiking based operation, in the device proposed by Berco et al.[13] as opposed to nonvolatile-state transitions. As for data processing, Hubel and Wiesel showed that mammalian neurons in the visual cortex are in fact stimulated by small regions of the visual field known as the Receptive Field.[14]

Ever since the introduction of the learning machine,[15,16] brain-inspired computing has strived to imitate the functionality of biologic systems. With regard to visual systems, artificial vision has to account for both the functionality of light-sensitive receptor cells and the analytical processing done by the visual cortex. Visual organs should therefore be the primary source of reference for a bioinspired vision system design. There are currently two main approaches to the implementation of bio-like machine vision. The first uses dedicated hardware centered around event-based vision sensors,[17–19] while the other is software-based and focuses on algorithmic execution.[20] Event-based vision sensors are utilized in special spiking cameras,[21] where control over the acquisition of visual information is transferred to unitary pixels that handles their information individually. This approach is quite different from conventional paradigms that rely on discrete time quantization in the form of pixel-array snapshots. As for software algorithms, these are mostly built around hierarchical feed-forward techniques and

---

DOI: 10.1002/aisy.202100025

© 2021 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH
commonly used for tasks such as object identification. In these approaches, a feed forward CNN implements two bracketed pairs of convolution operator followed by a pooling layer.[23] They are constructed according to the two visual pathways theory, stating that a primate visual cortex can be split between dorsal (where) and ventral (what) streams originating from the primary visual cortex.[22] However, they tend to be oversimplistic and are unable to successfully account for important aspects of perception such as detail preservation. Bioinspired, artificial smart retinas based on a 2D array paradigm of photoelectric logic gates were also suggested.[23–26] More recently, a three-input material nonimplication and logic conjunction (NIMPLY-AND) gate constructed out of two memristors,[27] and a pull-down resistor was demonstrated.[28] These gates may form building blocks for the design of a configurable array-matrix that effectively implements in situ image compression.

Herein, we present a system that was designed in an attempt to mimic the functionality of biologic visual systems while processing information contained in granular receptive fields by a CNN. First, the architectural concepts and algorithms are discussed, followed by an actual implementation using off-the-shelf components. This system was trained in two separate experiments to perform either pattern recognition or play Paper Scissors Stone (PSS) against a human player. Rotational invariance is demonstrated in the first experiment, while versatile, intelligent decision-making is shown in the second. Finally, the results of an online, supervised learning session are presented. This research can present an opportunity for the development of intelligent, visual perception apparatuses. Such systems could replace both bionic prostheses[29–31] and robotic eyes,[32] combining image sensors with artificial intelligent image processing into a single platform.

2. Results and Discussion

2.1. Architectural Concepts

The basis for image acquisition lies in the implementation of electro-optic, logical conjunction (AND) gates. These gates were arranged in the form of a 2D array. Visual information could then be directly imprinted onto this array as light inputs to corresponding gates to form a pixelated image. Figure 1a shows a schematic illustration of a single AND gate. It is constructed out of a light-dependent resistor (LDR) device S having one terminal connected to an electric input A and the other terminal (which also forms the output) to a passive resistor R. LDRs were chosen due to their transfer characteristics under illumination. Photodiodes usually show a sharp exponential transition in forward-bias current once the illumination threshold is reached (more compatible with representing a binary state transition). When reverse-biased, they display a linear behavior with the irradiation power intensity. Therefore, very large illumination variations would be required to generate equally proportionate current differences. LDRs, on the contrary, display a more gradual, exponential reduction in resistance as the light intensity increases. It is therefore more suitable for generating a wider range of analog levels as was used in this work. With LDRs, the overall sensitivity may therefore be tuned to respond to a much wider range of illumination intensity, while producing a gradual, Goldilocks-zone response, without saturating the gates (as seen in Figure 1).

The light input B is determined by the presence of illumination, and used to change the resistance of the LDR upon exposure. Both S and R form a current path from A to the common reference (Gnd) as the output voltage level $V_o$ is determined by a voltage divider. When the LDR is exposed to light, its resistance drops, and the resistance ratio $R/(R+S)$ changes. A change from $\approx 100 \text{k}\Omega$ in darkness to about 500–100 $\text{\Omega}$ under illumination ($10^3–10^4 \text{lux}$) was measured for S. The resistance R was implemented by using a potentiometer set to $\approx 1 \text{k}\Omega$. In this manner, as the gate was irradiated (B set to logic “1”) the analog output voltage level increased from $\delta$ to $\alpha$. Taking the resistance ratio into account, it corresponded to an output logic transition from “0” to “1.” This functionality is summarized in the truth table of Figure 1.

Biologic neurons are known to have an activation threshold. To account for a similar functionality in this work, a threshold level was determined based on the analog characteristics of
the gate. This threshold was used to define a state of being exposed to illumination for the system. It therefore allowed canceling-out environmental biasing throughout operation. Such a predefined threshold may also be adjusted during real-time operation to emulate relaxation sensory in living organisms. The threshold level was determined based on the simulation results shown in Figure 1c.

The simulation shows the output voltage \( V_o \) under different illumination intensities and supply levels (2.5, 3.0, and 3.3 V). Three operating conditions (low, typical, and high) were defined, based on the variance in LDR-irradiated resistance (±50%) and potentiometer tuning resolution (±10%). The figure therefore depicts three curves for each of the supply levels for a total of nine curves. The plots reveal that \( \delta \leq 0.1 \text{ V} \) in darkness and \( \alpha \approx 1.3-2.1 \text{ V} \) for levels above \( 10^3 \text{ lux} \). A threshold level of \( V_{TH} = 1 \text{ V} \) was thus defined (marked by a dashed line). This threshold is associated with an intensity that may vary from \( 10^2 \) to \( 5 \times 10^2 \text{ lux} \), depending on the LDR’s resistance variance (low, typical, and high operating corners). The inset in Figure 1c shows the circuit under simulation including the lumped capacitance across the output node. It should be mentioned that biological neuron activation thresholds are lower than the one defined in this work. Early models such as Hodgkin–Huxley describe the neuronal action potential as having amplitudes of up to 100 mV. \( V_{TH} \) in this work is determined primarily by the supply voltage \( V_{dd} \) that, in turn, is dictated by the requirements of the peripheral circuitry. This circuitry was implemented using off-the-shelf, 3.3 V components to demonstrate a prototype. Alternatively, low-power controllers or advanced application-specific circuits (e.g., 1.8 V and below) may be used to implement the same circuits. In such a case, the threshold level would be reduced by the same ratio because the AND gates are based on passive devices.

The AND gate may be regarded as an artificial neuron that operates on two weights (a variable voltage \( AV_{dd} \) and resistance \( R \)) and the light input to give an output subject to fulfilling a threshold condition. This enables a sensitivity parameter to be defined which may be used to determine when the system is exposed to illumination and thus to cancel out the environmental bias. For example, it can consist of a summation over the AND gates that cross the threshold and effectively flip to a high logic level. This number can be lowered to reflect a high-sensitivity or increased to reduce the sensitivity. Therefore, once a threshold is defined, the AND gates may be treated as digital entities as reflected by the truth table in Figure 1b.

The implementation of an image sensor capable of capturing pixelated image data implies efficient and compact placement of light-sensitive devices. A schematic diagram showing the system-level architecture is shown in Figure 2. An array of LDRs \( S_{ij} \) each having one terminal connected to an analog supply \( AV_{dd} \).
(effectively acting as the electric logic input A) is shown in Figure 2a. Exposing the array to light resulted in conductance changes in individual LDRs. In our chosen methodology, array information was evaluated on a row-by-row basis through the use of word-lines (WLs). For this purpose, selection devices \( M_{ij} \) were connected to individual LDRs and each row was activated by its corresponding WL. As the AND gate functionality requires an additional series resistor, four potentiometers \( R_i \) were placed at the termination points of the bit-lines (BLs). In this manner, activation of WL led to the formation of an analog voltage level \( V_{ij} \) on BL. These voltages were then buffered and converted to digital representations by a set of analog-to-digital converters (ADCs, marked A/D) and buffered by a controller. A set-vector was constructed once all the rows in the array were evaluated. This vector served as an input to a three-layer ANN.

Figure 2b shows the top-level conception of the dataflow. It shows the manner in which a row of optoelectric AND gates is evaluated. Their analog output levels are converted into digital format and input to the ANN. The computation result will be displayed to the user after all four rows are latched and the ANN output is valid. As part of this work, the ANN was trained to either perform pattern recognition or play PSS, and the result shown over a display module.

2.2. Artificial Vision System and Test-Bench Implementation

The concepts discussed in the previous section were implemented as a simplified prototype using off-the-shelf components. A schematic drawing showing the test-bench is given in Figure 3a. It consists of two parts that are the player’s move generator (shown on top) and vision system (bottom). A player may therefore initiate any random move in a game of PSS by pressing a corresponding soft-button on the screen of the Smartphone. The mobile device then communicates this move to a designated controller (CNT) through a Bluetooth channel. Once received, CNT manipulates the WLs and BLs of the LED array using a pre-determined sequence. It results in an image being generated over the LEDs and imprinted as resistance changes in the LDRs.

The images in Figure 3b,c depict the vision system prototype. In the figure, both the front and back sides of the system are shown, while either moved back (Figure 3b) or placed on top, in an aligned configuration with the LEDs (Figure 3c). Referring to Figure 3b, a \( 4 \times 4 \) array of LDRs (1), with an area of about \( 20 \times 25 \text{ mm}^2 \), and four BL-terminating potentiometers (2) were placed over the front face. As for the display (7), buffers (8), and selection devices (9), these were placed over the back side, as shown in Figure 3c. A microprocessor-based module (4) was programmed to implement the controller along with a three-layered, feed-forward, back-propagation CNN. The chosen configuration is considered as good, general purpose architecture for either supervised or unsupervised learning. WLs and the analog supply (3) were also driven by the controller, to allow acquisition of an image imprint over the LDRs. An \( 8 \times 8 \) light-emitting diode (LED) array (5), having an area of about \( 20 \times 20 \text{ mm}^2 \), was controlled through a gamepad (6) Bluetooth application running on an Android smartphone. These LEDs were driven by another separate and unrelated processor module. No communication channels were formed between the two processors except for LED–LDR interaction. Image patterns (i.e., Paper, Scissors, Stone, and All) were programmed into the second processor’s memory and activated by pressing the corresponding keypads on the application, as shown in Figure 3c. Each decision made by the ANN, being either pattern recognition or PSS response, was presented over the display module (7).

2.3. Testing Methodology

In an attempt to imitate the manner in which retinal ganglion cells spike in response to their conjugated bipolar cells’ polarization state, the LEDs were triggered on a row-by-row basis as well. It was done using a pulse width modulation (PWM) scheme that allowed control over their on-time through duty cycle variations. It is probably worth mentioning that the chosen approach is by no means exclusive, and other ways may be used to produce a similar outcome (e.g., PWM of the WLs’ in the LDR array instead). However, the overall result should be similar because,
as will be discussed later, the input set-vector values are determined by summation and root-mean-square (RMS) calculation of BL voltage levels. These RMS values determine whether the system is being exposed to illumination, by comparing them with the previously mentioned threshold level, much like the activation of biologic neurons.

Figure 4a,c,e,g,i contain images that show the different patterns generated using the LED array. Those patterns were designed and preprogrammed into the second controller. Each image is accompanied by a green-black 8 x 8 bit representation alongside it where green represents an on-LED and black an off one. Figure 4b,d,f,h,j shows the corresponding output for each image as produced by the ANN over the display module. The calculation outcome was defined to be either similar to the projected image (pattern recognition experiment) or a countermove to the player’s move in PSS. LED patterns were intentionally designed to inherently contain different levels of symmetry. Starting from Paper, that is asymmetric (Figure 4a), continuing with Scissors, having a single axis of symmetry (Figure 4e), and ending with Stone and All (Figure 4c,g,i) with multiple axes. Rotation-invariant functionality testing was performed while rotating the LDR array around its normal vector (z-axis). The purpose was to demonstrate that both multisymmetry shapes (Stone, All) and asymmetric (Paper) and y-symmetry (Scissors) could be identified by the CNN regardless of the relative LED–LDR arrays orientation.

2.4. Bioinspired Image Acquisition

This section details the bioinspired formulation and methodologies used to generate the input to the ANN. In biologic visual systems, off-bipolar cells display a hyperpolarization-like response under illumination that decays into a depolarized state in darkness. On the contrary, on-bipolar cells produce a depolarized output under illumination and hyperpolarize in darkness. This is an essential construction that allows for better accuracy and sensitivity in retinal center-surround cell configurations. The pixelated image was processed based on an attempt to mimic this biological behavior. Essentially, it was done through an exponential activation of the divergence of the gradient in discrete space (i.e., Laplace operator). Using the outcome of the second derivative over the array output effectively mimics the center-surround functionality. In this manner, the gradient between neighboring LDRs is amplified and fed into the ANN rather than the actual analog voltage level. This mimics event-based processing in biologic vision. The gradient may also be mapped to a binary number using a threshold level to further reduce the number of input nodes to the ANN. In the biologic counterpart, bipolar cells respond based on the exposure of the central cell, with reference to its neighboring cells as discussed previously. ANN inputs are thus generated by small regions in the visual field, in an attempt to account for Receptive Field stimulation.[14]

Rotation-invariant pattern recognition implies correct image classification while in a misaligned position, based on a supervised-training done using a dataset that corresponds only to an aligned placement. The approach chosen herein was to attempt generalization at the image acquisition stage while keeping the ANN simple. As mentioned previously, analog voltage levels produced by the AND gate configurations were read in a row-by-row basis, converted to a digital format and latched by the controller in a continuous manner throughout operation. These voltage levels were recorded over a predetermined number of sampling rounds, where each started with the activation of WL_0 and ended with WL_1, in a manner that latched the entire set of outputs V_o into a 4 x 4 matrix (A). One may think of this matrix as a potential over which the Laplacian operates to identify “sources” and “sinks.” Instead of feeding image data directly into a CNN, as done in conventional approaches, the sources-sinks map is being inputted. This abstraction allows for improved rotation-invariant
properties. The matrix $Q_{ij}$ was derived from these levels using the Arrhenius equation. A normalized activation parameter was generated by application of the Laplace operator with cyclic boundary conditions over this pixel analog information ($V_{o_{ij}}$), as detailed in Equation (1)

$$Q_{ij} = a \cdot \exp \left( \frac{V_{RMS}^{2}}{V_{AVG}} \Delta x \Delta y \right)$$

where $a$ is a unitless pre-exponential factor. $V_{RMS}^{2}_{ij}$ is the root-mean-square (RMS) value of $V_{o}$ over $n$ samples, $\Delta x = \Delta y = 1$, and $V_{AVG}$ is a moving average for all the measured outputs, over $m \geq n$ samples. $\Lambda_{RMS}^{ij}$ allows the evaluation of the energy embedded within a time-alternating signal and establishes an activation threshold as occurs in biologic neurons. The AND gate’s analog outputs $V_{o}$ thus facilitated the computation through latching and averaging of the 4 x 4 LDR array. $V_{AVG}$ is then calculated as follows

$$\Lambda_{RMS}^{ij} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} V_{o_{ij}}^{2}}$$

$$V_{av} = \frac{1}{i j} \sum_{i,j=1}^{4} V_{o_{ij}}$$

$$V_{AVG} = \frac{V_{av,m} + (m-1)V_{av,m-1}}{m}$$

$V_{av}$ is obtained by averaging over the entire analog input span during one sampling cycle. $Q$ may be viewed as a 1D vector $q_{i}$ with multiple rows reorganized into a single row. This vector was treated as a set-input to the ANN without any information loss as shown in Equation (5).

$$q_{i} = (Q_{1,1} \cdot Q_{4,1} \cdot Q_{1,2} \cdot Q_{4,2} \cdot Q_{1,3} \cdot Q_{4,3} \cdot Q_{1,4} \cdot Q_{4,4})$$

The ANN will remain in a standby mode until the array is exposed to a predetermined, minimal amount of irradiation. Once this limit is reached, the ANN is triggered to perform a calculation based on an analog input as detailed earlier. In this manner, the sensitivity of the system may be determined and

Figure 5. A winning streak in a game of PSS. The ANN can counter a human player’s with winning moves with a high success rate. a) Scissors in response to a player’s Paper. b) Stone in response to a player’s Scissors. c) Paper in response to a player’s Stone.

Figure 6. Rotational and spatial shifting of the relative placement between the LED and LDR arrays, as used throughout the experiment. a) Spatial shifting along the first axis (dy). b) Spatial shifting along the second axis (dx). c) Angular rotation of a positive angle (dt). d) Time evolution of the cumulative error after each training epoch for a training set consisting of 45 patterns. Each epoch consisted of ten iterations and lasted for about 4 s. A successful training was defined for an error lower than $10^{-3}$. Inset: Schematic depiction of the ANN, having two hidden layers and an output layer.
modified in real time to cancel out any environmental bias. When in standby, the state of the array is periodically evaluated to determine whether the ANN should be triggered. During this process, the output of each AND is compared with \( V_{TH} \) and a binary digit \( (W_{ij}) \) is assigned to the output. These bits are added and once the summation surpasses the sensory limit a trigger occurs. For \( n = 1 \), \( V_{RMS}^{ij} \) in a single sampling cycle can be used to determine whether the array is exposed to illumination while overriding environmental effects. An illumination condition can thus be defined based on \( \Lambda_{RMS}^{ij} \) and the simulation-based threshold \( V_{TH} \) multiplied by an empirical factor \( b \) (\( V^{-1} \))

\[
W_{ij} = 1 \text{ if } \frac{\Lambda_{RMS}^{ij}}{V_{AVG}} \geq b \cdot V_{TH}; \ 0 \text{ otherwise} \quad (6)
\]

The sensory limit is compared with a summation over all matrix elements

\[
\sum_{i,j} W_{ij} \geq \text{sensory limit} \quad (7)
\]

This approach may also be used to determine a saturation-dependent threshold and account for sensory relaxation found in living organisms’ retinas by resetting the average levels after a predetermined time period during which no illumination was applied.

**Figure 7.** Demonstration of rotation invariance in pattern recognition for an ANN trained using an input-set corresponding for a fully aligned state. Correct pattern identification with \( \approx 25^\circ \) of rotation: a) All. b) Paper. c) Scissors. d) Stone; correct pattern identification with \( \approx 12^\circ \) of rotation: e) All. f) Paper. g) Scissors. h) Stone; correct pattern identification with zero rotation (aligned): i) All. j) Paper. k) Scissors. l) Stone.
2.5. Paper Scissors Stone

The ANN was taught to correctly respond with a winning move in a game of PSS as shown in Figure 5. Training the ANN to either play PSS or perform pattern recognition essentially has the same level of complexity because the ANN’s output is simply interchanged in response to a certain input. The purpose of this experiment was to highlight the versatility and flexibility of the artificial visual system presented in this work. The system was trained offline, while during the experiment, it was able to counter the moves of a human player to a very high success rate (>95% for 50 rounds). The success rate may be improved even further by enlarging the size of the training set, as will be shown in the next section.

2.6. Rotation-Invariant Pattern Recognition

This section summarizes the results of an image classification experiment based on misaligned placements between the vision system and test-bench. Initially, the ANN was trained using a first input-set and up to a specific accuracy. The performance was then evaluated by counting the number of misidentified patterns from a series of random test images, for both angular and spatial misalignments. The analysis was then repeated for a second, larger training set, along with higher accuracy. It should be emphasized that the input training sets used throughout this experiment were strictly based on an aligned situation (zero displacement). Correct identification under misplaced conditions would therefore serve to indicate the successfulness of model generalization that accounts for rotational translations.

Figure 6a–c shows the definitions of misaligned placements. The figure shows skewed and tilted placements along the y, x and rotational axes, respectively, marked as dy, dx, and dt. For the first part of the experiment, a training set consisting of 45 input patterns was used. Figure 6d shows the commutative training error as a function of time. Training was arrested once the error dropped below $10^{-3}$. It took roughly 120 s to complete the entire session and the training consisted of about 35 epochs, containing ten iterations each.

Figure 7 shows successful image recognition under different rotations. Figure 7a–d shows correct identification of All, Paper, Scissors, and Stone, respectively for $dt \cong 25^\circ$. Figure 7e–h shows correct identification for $dt \cong 12^\circ$, and Figure 7i–l shows correct identification with $dt \cong 0^\circ$.

In the second part of this experiment, training was redone using a larger set containing 56 patterns. In addition, the exit-condition error was updated and reduced to $0.5 \times 10^{-3}$. The purpose was to evaluate the dependence of correct identification on training resolution and effort (i.e., larger sets imply larger efforts and a smaller error implies higher resolution). As shown in Figure 8a, the training took $\approx 350$ s (triple the effort). To evaluate the performance, 30 random images were tested in the first part and 50 in the second. A mismatch figure was calculated by taking the ratio of misidentified patterns over the total figure count. The resulting figures are shown in Figure 8b–d as a function of the displacements dx, dy, and dt.

These plots show that the mismatch in the aligned cases (dx = dy = dt = 0) was below 5% for both experimental parts. It can serve to indicate that the size of the training set used in the first part is indeed sufficient to yield a low bias. The error

![Figure 8](image-url)
bars in the plots correspond to a variance of $\pm 1$ in misidentification during the experiment. It is evident that increasing the set cardinality helped to reduce noise and improve performance especially for the rotational misalignment (Figure 8c) and $y$-shifting (Figure 8d). As far as misplacement along the $x$-axis, the larger cardinality helped to produce better results for a small displacement (2.5 mm point in Figure 8b). However, as the mismatch grew to 5 mm, the misidentification figure bounced back to over 50%. This should be regarded as an artifact caused by the physical construction of the LDR array prototype instead of an indication of a performance limit. It can be seen from Figure 3a that LDR devices were spaced roughly 5 mm apart in the $x$-direction (the board spacing is 2.54 mm and LDRs are placed 2-pitches apart). Once the LED–LDR overlapping field of vision was shifted by that amount, the LEDs were obscured from an entire column of LDRs, and the cyclic divergence calculation (Equation (1)) fell out of the generalization range of the model. On the contrary, in the $y$-direction the LDRs were still sufficiently exposed even with a 5 mm shift due to the smaller vertical spacing between LDRs.

2.7. Online Learning

The process of real-time, supervised learning implies a need to convey mistakes back to the ANN. Such errors may then be corrected on-the-fly, to tune and improve the overall performance. Here, we aimed to achieve classification by correctly labeling an increasingly growing training dataset while accounting for noisy inputs. Visual interaction was chosen as the main means to transmit feedbacks to the ANN. It was done so to keep the experimental methodology consistent and avoid a need for establishing new communication channels between the ANN and test-bench (independent controllers). This approach served to eliminate any doubts regarding the methodology in which the ANN corrected itself. In other words, the ANN has no indication as to the image generated by the test-bench controller apart from the visual error feedback. A human–machine handshake protocol was therefore defined in the following manner. Initially, the ANN was preset to recognize only the All pattern and the training set (input vector bank) kept empty. This pattern was then used to feedback mistakes to the ANN. Specifically, the shape All was used to indicate wrong responses, while the ANN was trained in real time to identify the other three patterns (PSS).

The ANN was thus trained to perform image recognition. This experiment progressed in stages, where at each stage a random pattern was displayed over the LEDs by a human player. This pattern triggered the generation of an input vector to the ANN as detailed by the previous sections. Each vector was either a new one or an already encountered one. For a new vector, a random response was produced and the vector was added to the training set along with the response. As for an existing vector, the ANN calculated the response based on its current state (i.e., weights). In either case, the human player then decided whether or not to indicate an error in his next move. If no error message was received, the state of the ANN was maintained. Once an error was indicated, the ANN acknowledged it, and retraining was initiated by labeling this vector with a different random response. Each training session was terminated as the training error fell below $10^{-3}$ arbitrary units.

Keeping these constraints in mind, no matter which approach was chosen to indicate errors, it is essential that the learning process converges to a successful outcome. For this purpose, a success rate (0–1) was defined and monitored throughout the training process. The reader is advised to consult the supplementary information for more details. Once this rate stabilized near 1, the ANN was considered to be trained. The evolution of the learning process is shown in Figure 9 along with the success rate and timing characteristics. It is evident from the figure that the ANN successfully reached a trained state with high success after $\approx 800$ s.

![Figure 9](link).

**Figure 9.** Temporal evolution of the online learning process. a) Success rate and new pattern count as a function of time. The success rate fluctuated as new patterns were encountered. However, it eventually rebounded as learning progressed and converged close to 100%. Total pattern count in the training set grows from zero to 16 as the learning process evolves. b) Total training time for each new session invoked by an error feedback during the learning process. c) Training time measured for each new session.
As mentioned earlier, the training set was empty at the beginning of the experiment. Therefore, as shown in Figure 9a, the success rate was very low because virtually all the encountered vectors were new to the ANN. Once training progressed and the pattern count increased, the success rate increased and fluctuated. Bumps were caused by the appearance of new patterns and mistaken decisions made by the ANN. However, the system was able to eventually converge to a high rate, indicating successful training. Further support for this conclusion is given by the overall training time shown in Figure 9b and the training time for each pattern in Figure 9c. The bars in the plots indicate the measured time it took to complete each retraining session. Contrary to the notion that the training process would be lengthened as more patterns were successively added into the set, the overall time decreased along with the time per pattern. This helps to support the assumption that a converging learning process took place.

3. Conclusions

In summary, this work presented an artificial intelligent, bioinspired robotic vision system with rotation-invariant properties. The concept was demonstrated by off-the-shelf components, used to fabricate an ANN-based system that was taught both offline and in real time to perform pattern recognition tasks. The architecture was based on optoelectronic AND gates, treated as building blocks to implement a sensory array. Analog voltage levels, produced by the said gates in response to an illuminated pattern, are directly correlated to the pixel information of this image. In this manner, image data were collected in a row-by-row basis to construct an input set-vector to an ANN, which, in turn, calculated a decision according to predetermined criteria.

First, the architecture and operation principals were discussed along with the experimental setup and test procedures. It was then followed by a functionality demonstration and performance analysis for a game of PSS. During this experiment, the ANN was trained to successfully identify patterns projected over the vision sensor by a white-light LED array construction. These projected images were based on moves made by a human player. The system then countered those plays with a move of its own, with a high success rate. Tolerance to spatial translation and rotation was studied as well after training the ANN with a dataset that corresponded to an aligned orientation. Misidentification error during pattern recognition was then characterized for various degrees of misalignments. Finally, the process of online supervised learning was shown through real-time pattern recognition. The concepts presented in this work could help pave the way toward implementation of light-responsive gate arrays for futuristic bioinspired, intelligent vision platforms.

4. Experimental Section

Image pattern generation was done over an \(8 \times 8\) white LED array, each with an irradiation power density of about 20 mW cm\(^{-2}\). The array was operated by selecting designated BLs and cycling the rows to a high logic level. WL and BL were driven by the pads of an ESP32 control module. Player instructions were supplied by the user to the controller over Bluetooth communication using an Android gamepad application. Each button on the gamepad was therefore configured to produce a different image over the LED array.

The \(4 \times 4\) light-sensor array was fabricated with off-the-shelf light-dependent resistors, each connected to a VN0104 selection device. The array’s bit lines were terminated with four 3296 potentiometers, each set to a resistance of \(\approx 1\,\Omega\), to form a logic AND function. Analog voltage levels at the gates’ outputs were buffered through an LM324 differential amplifiers integrated circuit and converted to digital levels by four ADC channels from a second ESP32 module. The deep layers of the ANN and system controller were implemented over this module as well. ANN decisions were displayed on a 0.91 inch, 128 \( \times \) 32 monochrome unit over i2C communication protocol.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge the partial funding support by Singapore Ministry of Education under grants MOE2016-T2-1-102 and MOE2016-T2-2-102. Note: The affiliations were corrected on 23 August 2021, after initial publication online.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

D.B. conceived the concept, developed the architectural methodologies, implemented the experimental setup and modeling, performed the experiments and simulations, created the visualization, and coauthored the manuscript. D.S.A. acquired funding, provided technological insights and editorial guidelines, supervised the work, and coauthored the manuscript.

Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

Keywords

bioinspired machine vision, cognitive artificial retinas, electrophotonic logic computation, optoelectronic logic gates, robotic vision

Received: February 17, 2021
Revised: April 9, 2021
Published online: June 2, 2021

[1] F. Rosenblatt, in The Perceptron-A Perceiving and Recognizing Automaton, Cornell Aeronautical Laboratory, Buffalo, NY 1957.
[2] F. Rosenblatt, Psychol. Rev. 1958, 65, 386.
[3] W. Zhang, in Proc. of Annual Conf. of the Japan Society of Applied Physics, Tokyo, Japan 1988.
[4] W. Zhang, K. Itoh, J. Tanida, Y. Ichikawa, Appl. Opt. 1990, 29,4790.
[5] S. Dieleman, K. W. Willett, J. Dambre, Monthly Notic. R. Astronom. Soc. 2015, 450, 1441.
[6] X. Shen, X. Tian, A. He, S. Sun, D. Tao, in MM’16: ACM Multimedia Conf., Amsterdam, The Netherlands, Association for Computing Machinery, New York, NY, USA, October 2016, pp. 1345–1354.
[7] G. Cheng, P. Zhou, J. Han, IEEE Trans. Geosci. Remote Sens. 2016, 54, 7405.
[8] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, Y. Yin, IEEE J. Biomed. Health Inform. 2017, 22, 184.
[9] G. Cheng, J. Han, P. Zhou, D. Xu, IEEE Trans. Image Process. 2018, 28, 265.
[10] S. H. DeVries, Neuron 2000, 28, 847.
[11] H. Wässle, B. B. Boycott, Physiol. Rev. 1991, 71, 447.
[12] R. H. Masland, Nat. Neurosci. 2001, 4, 877.
[13] D. Berco, D. S. Ang, H. Z. Zhang, Adv. Intell. Syst. 2019, 1, 1900115.
[14] D. H. Hubel, T. N. Wiesel, J. Physiol. 1959, 148, 574.
[15] A. M. Turing, Proc. London Math. Soc. 1937, s2–42, 230.
[16] M. Turing, Mind 1950, LIX, 433.
[17] P. Lichtsteiner, C. Posch, T. A. Delbruck, IEEE J. Solid-State Circuits 2008, 43, 566.
[18] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, T. Delbruck, Proc. IEEE 2014, 102, 1470.
[19] L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, Nature 2020, 579, 62.
[20] N. K. Medathati, H. Neumann, G. S. Masson, P. Kornprobst, Comput. Vision Image Understand. 2016, 150, 1.
[21] D. Cox, T. Dean, Curr. Biol. 2014, 24, 921.
[22] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez, L. Wiskott, IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1847.
[23] D. Berco, D. S. Ang, Adv. Intell. Syst. 2019, 1, 1900003.
[24] D. Berco, D. S. Ang, J. Phys. D 2020, 53, 25LT02.
[25] D. Berco, Y. Zhou, S. R. Gollu, P. S. Kalaga, A. Kole, M. Hassan, D. S. Ang, ACS Nano 2018, 12, 5946.
[26] D. Berco, D. S. Ang, J. Vac. Sci. Technol. B 2019, 37, 061806.
[27] L. O. Chua, IEEE Trans. Circuit Theory 1971, 18, 507.
[28] D. Berco, D. S. Ang, P. S. Kalaga, Adv. Intell. Syst. 2020, 2, 2000079.
[29] K. Mathieson, J. Loudin, G. Goetz, P. Huie, L. Wang, T. I. Kamins, L. Galambos, R. Smith, J. S. Harris, A. Sher, D. Palanker, Nat. Photonics 2012, 6, 391.
[30] Y. H. Luo, L. Da Cruz, Prog. Retinal Eye Res. 2016, 50, 89.
[31] J. Lee, C. Choi, D. H. Kim, Y. M. Song, Adv. Funct. Mater. 2018, 28, 1705202.
[32] B. K. P. Horn, Robot Vision, MIT Press, Cambridge, MA 1986.
[33] S. H. DeVries, Neuron 2000, 28, 847.
[34] A. L. Hodgkin, A. F. Huxley, B. Katz, J. Physiol. 1952, 116, 424.
[35] L. Hodgkin, A. F. Huxley, J. Physiol. 1952, 117, 500.