Human binocular color fusion model based on BP Neural Networks prediction

Yuxiang Zhu

Chongqing Yucai Secondary School, Chongqing,400000,China

Abstract. Stereoscopic display vision is significantly impacted by the color distortion of left and right eye images. When the human eye receives a specific range of dissimilar color information separately, the visual system combines them into a single color through binocular color fusion. In this study, we present experimental findings which compare the accuracy of a common binocular color-fusion model that was trained utilizing both linear fitting and back-propagation neural networks. Patient binocular color contrast test data was collected by eye care professionals working in private eye clinics. The results indicated that the back-propagation neural network produced RMSE errors of 0.9819 and 0.9662 for predicting binocular contrast, which were superior to the linear fitting method with errors of approximately 0.5. The BP neural network algorithm employed demonstrates predictive capabilities and lessens the occurrence of color redundancy. This reduction in redundancy holds the potential to decrease expenses associated with stereo imaging in future applications.

Keywords: BP neural network, binocular color fusion, stereoscopic display.

1. Introduction

The 21st century heralds the age of information technology, with the proliferation of computer technology and the internet leading to the popularization of digital media. Traditional book and voice information are being gradually displaced by images and videos. With the fast-paced advancements in display technology, three-dimensional (3D) stereoscopic displays are increasingly being utilized within the video game entertainment, visual research, remote device operation, medical imaging, virtual reality (VR), augmented reality (AR) and many other fields [1, 2] . Stereoscopic display technology is founded upon the stereoscopic vision of the human eye[3]. The visual system of the human eye generates a sense of stereoscopy through binocular and monocular, as well as physiological and psychological cues[4]. Physiological aspects of the human eyes, such as convergence, focusing, and binocular parallax, are the crucial stereoscopic cues for depth perception. Both binocular parallax and visual fusion also contribute to this perception[5].

Three-dimensional stereoscopic perception is primarily achieved through the parallax between the image pairs viewed by the left and right eyes. The human visual system employs temporal and spatial interlacing to segment the image, resulting in parallax. This is currently the most widely used method for stereoscopic display. When the color information presented to both eyes in a stereoscopic display is inconsistent, simultaneous stereo fusion and color fusion will occur, which increased information processing load on the visual system and predominantly leading to visual fatigue[6].

When both eyes receive colors within a specific range, those colors are fused by the visual center into a stable perceptual color, this phenomenon is referred as binocular color fusion. Most individuals possess two ordinary eyes, yet we perceive the visual world as a unified whole to a certain degree. Usually, our brains receive and analyze two distinct and mainly unconnected visual stimuli from the external environment[7]. If the colors perceived by the left and right eyes are dissimilar and inconsistent, it can result in color competition, suppression, or overlap, leading to significant discomfort. By contrast, when two colors are fused by the human eye, the resultant perceived color is typically intermediate between the two. Consequently, visual cues for fusion may include the relative similarity in color between the eyes. However, the occurrence of disparate colors in each eye is a contentious topic in two-color blending. It remains a matter of debate whether the amalgamation of colors of light perceived by each eye that present different hues is equivalent to the outcome when they are overlaid on a single eye.

In the early 18th century, researchers utilized various colors of silk to observe through openings. In the following century, Helmholtz and Herring disagreed over the occurrence of two-color mixing[8], as their color vision hypotheses received different predictions. One of the most significant findings from Anstis and Rogers' study was that "two eyes are better than one." The research data indicated that combining two colors will results in losing the visible features of one eye with a different color[7]. One of the most well-debated issues was whether "binocular yellow light" is produced when one eye received red light and the other green light. Wight and Wade confirmed the validity of these rule, and their study demonstrated that the combination of the solid color discs on the top resulted in an unsteady binocular rivalry, whereas the combination of the two textured discs on the bottom leads to a consistent bicolor blend[9]. This demonstrates how texture can assist in blending two colors.

Numerous scholars have analyzed the impact of images on display with regards to human visual characteristics[10]. They have also explored the influence of asymmetrical color information on the visual comfort of stereoscopic displays from varying viewpoints, namely luminance, hue, and chromaticity. Their research offered guidance for the development of stereoscopic image coding schemes. Dominic and other researchers proposed an experimental image to study the relationship between color and depth perception[11]. The experimental results indicate that color and depth are processed independently in the visual system, and their interaction occurs at a higher level of visual perception. Both the investigation of the relationship between color fusion and stereo fusion in the visual system and the practical exploration of using color to enhance the comfort of stereoscopic displays necessitate quantitative experimental data on the psychophysics of vision[12]. Moreover, the experiments demand meticulous design of the stimulus images and consideration of various parameters. Additionally, a significant amount of exploration of the topic is required.

In this paper, we refer to the visual psychophysics of binocular color fusion experiments. Measured data obtained from binocular fusion experiments, where different colors were induced on stereoscopic displays by medical professionals in our private eye clinic, serves as our basis. We utilized a gradient descent technique founded on the Levenberg-Marquardt transfer function to instruct a back propagation (BP) neural network. Through this approach, we constructed a prediction model that fuses binocular colors utilizing pre-existing measurements. In this paper, the available experimental data was processed and analyzed to establish both a mathematical and neural network model for predicting binocular color fusion. The neural network binocular prediction method yielded lower RMSE errors, 0.9819 and 0.9662 respectively, compared to linear fitting. This approach effectively reduced color redundancy, resulting in cost savings in stereo imaging for possible future applications.

2. Binocular color fusion model

When colors are received by the left and right eyes inconsistently, color competition, suppression or overlap will occur when color variation appeared, and the human eye will feel uncomfortable at this time13. The human eye is capable of fusing two colors into one and perceiving a color that lies between them. This phenomenon is explained by the binocular guidance model in Figure 1. The model can be comprised into two stages. Initially, binocular color sensing is triggered by a feedback mechanism that diminishes the color difference between two monocular images. This significantly impacted the color perception of monocular objects in the eye. Subsequently, binocular fusion converts the two monocularly induced colors into the ultimate fused color.

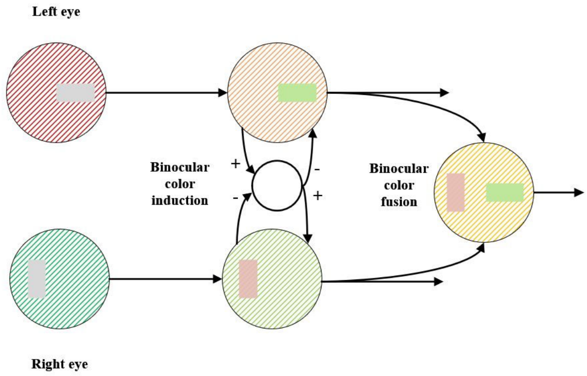

Fig. 1. Schematic of binocular color induction model

The appearance of two-color mixing relies on the brain's interpretation of information from both eyes as information derived from the same object, also known as the "object commonality hypothesis.". Among the models to simulate binocular fusion, the most straightforward approach is the weighted average model proposed by Willem[14]. When the left eye shows a color block that is uniformly colored and has a brightness of L_L, and the right eye shows a color block with a brightness of L_R, the resulting fused color block can be matched by calculating a weighted average.

$$L_{\text{fuse}}=aL_L+(1-a)L_R \tag{1}$$

The stable fusion of two colors into a single color, known as static perception, is the model of interest. This model is based on Legge and Foley's theory[15], and the resultant fusion can be expressed using the following model.

$$R=\frac{C^p}{\mathrm{z}+C^q} \tag{2}$$

In the formula, R is the contrast, and the relationship between contrast and C is internal response; C represent the images received by eyes. z, p and q received by the eyes are all constants, p, q>0 respectively, and are between 2 and 5, $p > q$; In this model, the contrast C is transformed into an acceleration function near the threshold, and then q and p can be determined under different contrast, and the contrast can be obtained by solving z.

Considering the mutual inhibition of eyes, Equation 4 can differentiate between the colors of eyes, represented by C_L and C_R, without subjective evaluations. The equation models contrast transformation as a function of acceleration near threshold and deceleration at high contrast to obtain a binocular contrast model, specifically a late summation model. Subsequently, the following equation is derived：

$$R(C_L,C_R)=\frac{C_L^P+C_R^P}{z+C_L^q+C_R^q} \tag{3}$$

Also, upper formula can be written in the following pattern, where a_0, b_0 is constant separately.

$$R(C_L,C_R)=a_0C_L+b_0C_R+C \tag{4}$$

In certain models, the exponent q may be excluded or the binocular response's nonlinearity may be enhanced. Assuming that a weighting function ω is incorporated into the interocular inhibition component to portray the degree of stimulation by the other eye, which characterizes the level of responsiveness and matching, the binocular response model can be expanded to the formula below.：

$$R(L,R)=\frac{C_L^P}{z+C_L^q+\omega C_R^q}+\frac{C_R^P}{z+C_R^q+\omega C_L^q} \tag{5}$$

Since the effect of high-contrast stimuli on the fused images was greater than predicted by simple linear summation, we introduced an interocular contrast gain function for the C_L and C_R images for each eye.

$$R(L,R)=\frac{C_L}{1+\varepsilon_R(C_R)}+\frac{C_R}{1+\varepsilon_L(C_L)} \tag{6}$$

Where, $\varepsilon_R(C_R)$ and $\varepsilon_L(C_L)$ are the total visually weighted contrast energies of the two input image gain controls. While the process of gain control remains the same, the computational paths used to perceive image phase, contrast, and depth information differ. The theory precisely describes the physiological consequences of binocular fusion and navigates the binocular color fusion to anticipate the fusion result. Consequently, training of color fusion visual data on a small batch of datasets is required to adequately fit the model. BP neural networks can be implemented to address this issue. Doctors from our private ophthalmology clinic have supplied a number of color contrast test samples of patients' LCD displays. This paper validates the model with the aid of these samples.

3. Experiment

3.1 BP neural network

A multilayer feedforward neural network trained using the error backpropagation algorithm is known as a backpropagation (BP) neural network. The technical term abbreviations will be explained upon first usage. In figure 2, the Backpropagation (BP) network incorporated multiple concealed layers between inputs and outputs, and alterations in the nodes' states in this layer have a direct impact on the connection between the inputs and outputs. The BP network's computational process is classified as forward and backward. Inaccurate output data from forward step-by-step processing will cause the signal to revert, and weight changes in each node will alleviate the error.

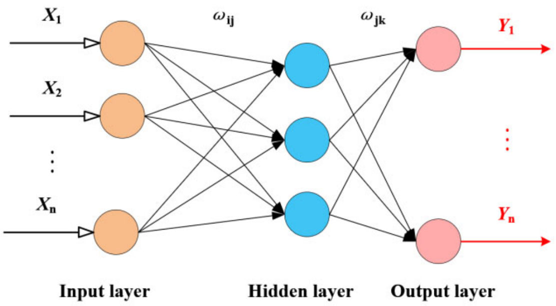

Fig. 2. Topological structure diagram of BP neural network

A backpropagation (BP) neural network can be viewed as in figure 2, X_1, X_2 … X_n represent the input stimuli for the network and the independent variables for the corresponding function. The values Y_1 … Y_n represent the network's projected output stimuli and serve as the dependent variable for the corresponding function.

The weight of the BP neural network is represented by the constants ω_{ij} and ω_{jk}. The network interprets the relationship between n independent variables and m dependent variables using the function mapping. To predict with BP neural networks, the crucial factor is to train the network by establishing the function mapping relationship, which involved multiple steps as follow.

Step 1: Initialization of the neural network. The number of nodes in the input layer, hidden layer l, and output layer m are determined based on the input-output sequence (X ,Y) of the system. Additionally, the connection weights ω_{ij} and ω_{jk} between the neurons of the input, hidden, and output layers are initialized. The thresholds for layer l and output layer m are also initialized with the learning rate and neuron excitation functions given by upper layer variables.

Step 2: Calculation of Hidden Layer Output. The hidden layer output H is calculated instantly based on input variable X, connection weight ω_{ij} linking the input layer and the hidden layer, and the threshold a for the hidden layer, depicted as follow.

$$H_{\mathrm{j}} = f\left(\sum_{i=1}^{n}\omega_{ij}x_i - a_i\right) \qquad \mathrm{j}=1,2,\mathrm{K}\ ,l \qquad (7)$$

Where l represents the number of nodes in the hidden layer, and f represents the hidden layer excitation function, which can be expressed in different manners. The function presented in this article is articulated as follows:

$$f\left(x\right) = \frac{2}{1-e^{-2x}} - 1 \qquad (8)$$

Step 3: Calculate the output of the output layer. The predicted output O is computed based on the implicit layer output H, connection weights ω_{jk}, and threshold b. The following formula expresses the computation.

$$O_j = \sum_{j=1}^{n} H_j\omega_{jk} - b_k \qquad k = 1,2,\mathrm{K}\ ,m \qquad (9)$$

Step 4: Error Calculation. The network prediction error e is calculated by comparing the predicted output O to the expected output Y. Both outputs are provided by the network.

$$e_k = Y_k - O_k \qquad k = 1,2,\mathrm{K}\ ,m \qquad (10)$$

Step 5: Weight updates. The weights of the network connectionsω_{ij} and ω_{jk} are adjusted based on the prediction error e. The learning rate η in formula determines the magnitude of the update.

$$\omega_{ij} = \omega_{ij} + \eta H_j\left(1-H_j\right)x\left(i\right)\sum_{i=1}^{n}\omega_{jk}e_k \qquad i=1,2,\mathrm{K}\ ,n;\quad j=1,2,\mathrm{K}\ ,l \qquad (11)$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \qquad j=1,2,\mathrm{K}\ ,l;\quad k=1,2,\mathrm{K}\ ,m \qquad (12)$$

Step 6: Update the threshold. Update the thresholds of the network nodes a and b, respectively, based on the error in network prediction, e.

$$a_j = a_j + \eta H_j\left(1-H_j\right)\sum_{i=1}^{n}\omega_{jk}e_k \qquad j=1,2,\mathrm{K}\ ,l \qquad (13)$$

$$b_k = b_k - e_k \qquad k=1,2,\mathrm{K}\ ,m \qquad (14)$$

Step 7: Judge whether the algorithm iteration is complete, and if not, return to step 2.

The process for the binocular color fusion system can be viewed as a black box. Firstly, a BP neural network is trained using measured data to derive an expression for binocular color fusion. This trained network is then used to predict binocular color fusion results, which are subsequently validated. Figure 3 provides a detailed flowchart of the binocular color fusion model fitting algorithm based on BP neural network.

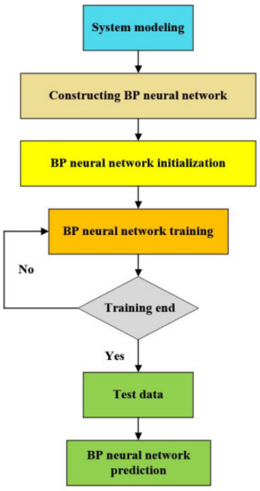
System modeling
Constructing BP neural network
BP neural network initialization
BP neural network training
No
Training end
Yes
Test data
BP neural network prediction

Fig. 3. Flow chart of BP neural network algorithm

The number of nodes in the hidden layer significantly affects the network's prediction accuracy. If the number of nodes is too low, the network cannot perform the study effectively, necessitating an increase in training times. Additionally, accuracy of training suffers. Conversely, with excessive nodes, training time will increase and the network is more prone to overfitting.

$$l < \sqrt{(m+n)} + a \tag{15}$$

In this study, the optimal number of hidden layer nodes is determined through a combination of formula and trial and error. Specifically, we apply the selection formula to identify the ideal number of hidden layer nodes. Then, we calculate this number utilizing the trial and error method. The formula utilized n for the input layer nodes, l for the hidden layer nodes, m for the output layer nodes, and a as a constant value between 0 and 10.

3.2 Training result

Previous experiments utilized a Samsung 3D monitor to exhibit color stimuli. The monitor features a 2D/3D switching capability with accompanying glasses needed for 3D display. Its physical dimensions measure 510 mm × 280 mm (horizontal × vertical), while its resolution amounts to 1920 (horizontal) × 1080 (vertical) pixels. It operates at a refresh rate of 144 Hz and was connected to a computer. The subjects in the experiment were one person, and two sets of experiments were conducted. The doctor at our private eye clinic collected the samples during a previous LCD color test for a patient.

Based on the properties of the binocular color fusion model, we designed the architecture of the BP neural network. The function includes two inputs, x and y chromaticities and one output chromaticity, $f(x, y)$, while each chromaticity is determined by the values of C_R and C_L. The BP neural network processed a 4-10-2 configuration with an input layer of four nodes, a hidden layer of ten nodes, and an output layer of two nodes.

In previous experiments, 320 sets of binocular color fusion data were obtained. Of these, 300 sets were adopted for training the network to predict binocular color fusion results, while the remaining 20 sets were employed for testing the prediction capability of the network. In this paper, the Levenberg-Marquardt algorithm is utilized to train the network, which combined the advantages of gradient method and Newton's method and is widely used.

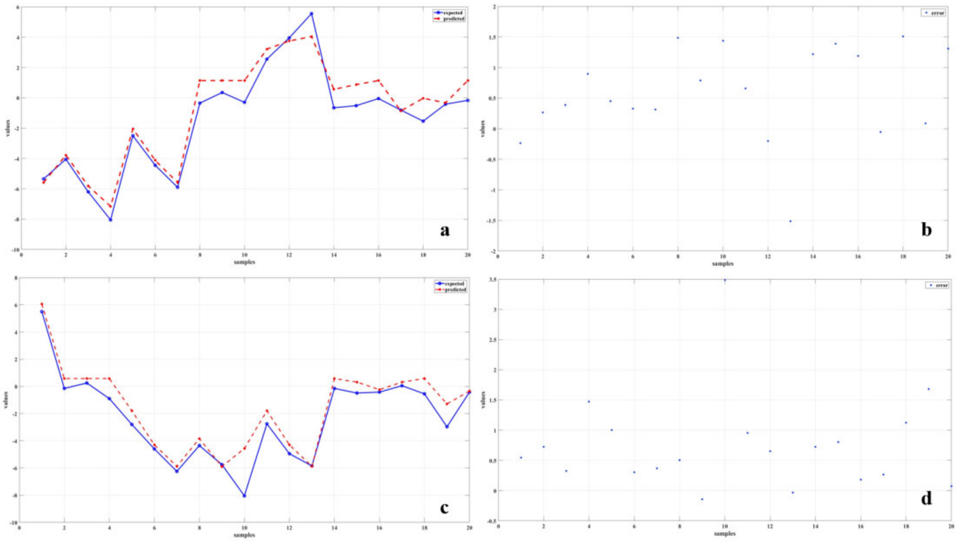

Fig. 4. BP best training result for binocular color fusion model (20 predicted samples), (a)(c)figure represent left and right eye training prediction respectively, the blue line is expected output result acquired from samples, red dot line is predicted output result trained from BP neural network algorithm;(b)(d) figure represent left and right eye training error from samples respectively, shown as blue dot.

The hyperbolic tangent S-shaped transfer function was selected as the excitation function for the hidden layer. To test the accuracy of the neural network prediction, the predicted output color is compared to the output color of the original measurement. Figure 4 displayed the expected and predicted output color values. The expected output represents the color value obtained from the sample, while the predicted output reflects the color value predicted by the model.

RMSE represents Root Mean Square Error, SSE is the sum of squares of the errors, the closer the SSE is to 0, the better the model fits and the more successful the data prediction received. The trained BP neural network has a certain prediction ability, and the RMSE values of the left eye and right eye are 0.9819 and 0.9662, respectively. The closer the RMSE value is to 1, the better the model is and the higher the prediction accuracy. The RMSE of the left eye in the figure is greater than 0.98, indicating a better fit of the neural network. The prediction of the right eye appears to be worse than that of the left eye because there are many singularities in the right eye data.

3.3 Training comparation

First, we fit the averaged data with a model to create a quantitative equation for binocular color fusion perception. Based on the conclusions of the data analysis described above, we selected a weighted average model for fitting. Here, x and y are the input variables of the left and right eyes, which can be considered as CL and CR; a is a constant, approximately equal to 0.49, which needs to be obtained by fitting. The linear fitting results are shown in figure 5(a).

$$f(x,y)=ax+(1-a)y \tag{16}$$

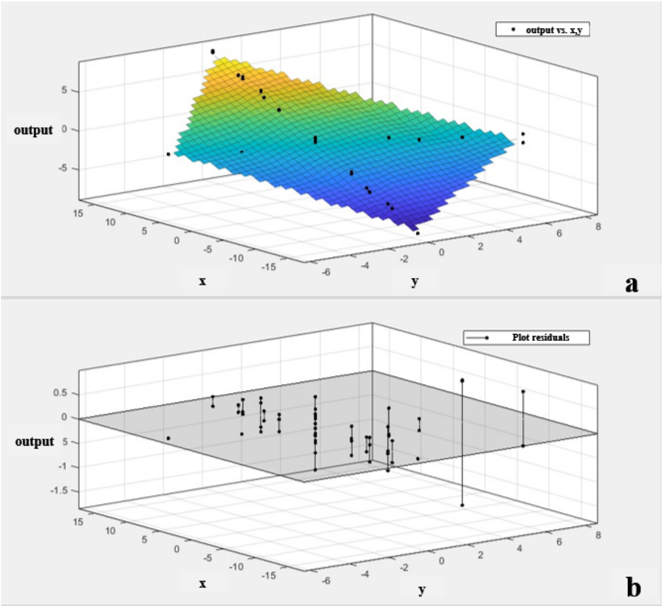

Fig. 5. Typical linear fitting results as f(x,y)=ax+(1-a)y ,which vertical axis output indicate as f(x,y);(a)curve fitting using output data.(b)residuals for samples to show fitting error.

The comparison of the fitting results is shown in table 1.R-squared is a deterministic coefficient that takes values between 0 and 1. The closer it is to 1, the greater the explanatory power of the input variables in the equation and the better the fit of the model. Adjustment R-squared is the adjustment and determination of the coefficients. When variables are added, even meaningless ones, the R-squared remains the same or increases. The adjusted R-squared penalizes added variables that do not contribute to the effect of the model. As more meaningless variables are added, the gap between the R-squared and the adjusted R-squared will increase. However, if the added eigenvalues are significant, the adjusted R-squared will also increase.

To obtain a better model, we fit the weighted average model and added the parameter b. The fit results are shown in figure 6. So far, the binocular color fusion model can be described as follows.

$$f(x,y)=ax+(1-a)y+b \tag{17}$$

As shown in Table 1, both R-squared and Adjusted R-squared are increased and both RMSE and SSE are decreased compared to the model without compensation, indicating a better fit. The R-squared and Adjusted R-squared are not significant for the BP neural network. In addition, the RMSE, which is our main concern in linear fitting, is about 0.5, which is much lower than that of the BP neural network.

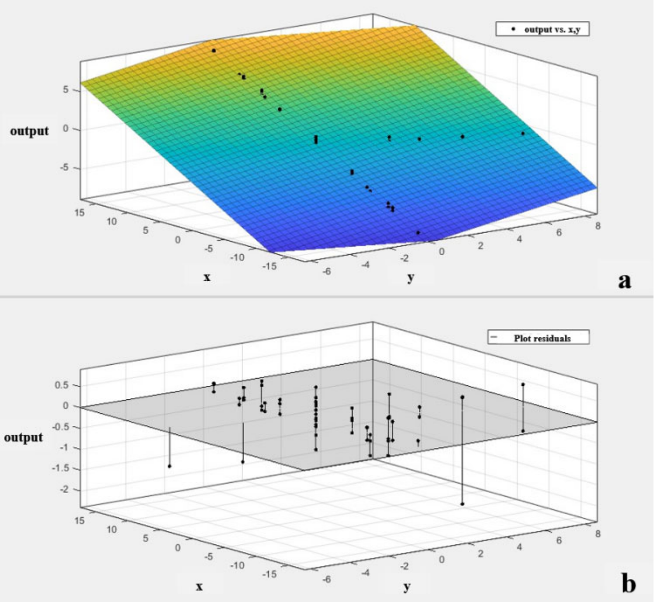

Fig. 5. Typical linear fitting results as f(x,y)=ax +(1-a)y+b with compensation term ,which vertical axis output indicate as f(x,y);(a)curve fitting using output data v.s. (b)residuals for samples to show fitting error

Table 1. Comparation of linear fitted and BP training

Fitted function	RMSE	SSE	R-square	Adjusted R-square
f(x, y) =ax+(1-a)y	0.5586	19.72	0.933	1.042
f(x, y) =ax+(1-a)y+b	0.5342	17.7	1.682	0.981
BP(left eye)	0.9819	2.276	-	-
BP(right eye)	0.9662	2.488	-	-

These experimental results show that the predictions obtained from the neural network trained model are closer to the experimental data than the linear fit.

4. Conclusion

In this study, we trained a Levenberg-Marquardt algorithm BP neural network applying measurement data from binocular fusion experiments on stereoscopic displays with varying color contrasts conducted by doctors in our private eye clinic. Our main objective was to build a prediction model for binocular color fusion based on the results of these experiments in the context of visual psychophysics research. A prediction model for binocular color fusion was established through the implementation of a Levenberg-Marquardt algorithm BP neural network. The available experimental data was processed and analyzed to establish the mathematical and neural network models for predicting binocular color fusion. The RMSE errors of the binocular prediction neural network are 0.9819 and 0.9662, respectively. These values are superior to those obtained by the linear fitting technique. The training outcomes of the BP neural network approach are more precise compared to the typical linear fitting method. We employed a method that reduces color redundant information, thus resulting in a decrease in stereo imaging costs for future applications. Limited by the number of training samples, higher training accuracy and model fitting accuracy could potentially be achieved with the inclusion of additional samples and training sets. However, achieving large-scale training with tens of thousands of samples requires further work.

References

[1] F. G. Segala, A. Bruno, J. T. Martin, M. T. Aung, A. R. Wade and D. H. Baker, Elife 12, RP87048 (2023).

[2] J. Shen, Y. Zhang, Z. Liang, C. Liu, H. Sun, X. Hao, J. Liu, J. Yang and L. Shao, IEEE Transactions on Circuits and Systems for Video Technology 28 (5), 1158-1168 (2016).

[3] J. Skerswetat and P. J. Bex, Consciousness and Cognition 107, 103437 (2023).

[4] W. J. Tam, F. Speranza, S. Yano, K. Shimono and H. Ono, IEEE transactions on broadcasting 57 (2), 335-346 (2011).

[5] S. Moussaoui, C. F. Pereira and M. Niemeier, Cortex 159, 26-38 (2023).

[6] L. Libenson, Journal of Vision 15 (5), 2-2 (2015).

[7] S. Anstis and B. Rogers, Vision research 53 (1), 47-53 (2012).

[8] J. T. Desaguliers, Philosophical Transactions of the Royal Society of London 29 (348), 448-452 (1716).

[9] H. R. Wilson, R. Blake and S.-H. Lee, Nature 412 (6850), 907-910 (2001).

[10] H. Komatsu, Current opinion in neurobiology 8 (4), 503-508 (1998).

[11] K. R. Gegenfurtner, D. C. Kiper and J. B. Levitt, Journal of neurophysiology 77 (4), 1906-1923 (1997).

[12] Z. Zhang, C. Han, S. He, X. Liu, H. Zhu, X. Hu and T.-T. Wong, The Visual Computer 35, 997-1011 (2019).

[13] C. J. Erkelens and R. van Ee, Vision Research 42 (9), 1103-1112 (2002).

[14] D. R. Simmons, Perception 34 (8), 1035-1042 (2005).

[15] G. E. Legge and J. M. Foley, Josa 70 (12), 1458-1471 (1980).