Recognition of Artificial Ripening Tomato and Nature Mature Tomato Based on the Double Parallel Genetic Neural Network

1, 2Haibo Zhao and 3Xianghong Zhou
1Engineering Technology Research Center of Optoelectronic Appliance, Anhui Province,
2Department of Electrical Engineering, Tongling University, Tongling Anhui, 244000, China
3No. 43 Research Institute, China Electronic Science and Technology Group
Company, Hefei Anhui 230088, China

Abstract: In order to prevent artificial ripening tomato into markets to harm consumers' health, a double parallel genetic neural network identification system was designed. This system obtained tomato external color characteristic parameters (R, G, B) through the computer vision device and changed the RGB value into HIS value. Put tomato external color characteristic parameters as input, tomato maturity properties as output and verified the system with test samples. The test results show that, the correct recognition rate of the system is 93.8%, providing the reference for further research of artificial ripening tomato and natural mature tomato.

Keywords: Artificial ripening tomato, genetic algorithm, natural mature tomato, neural network

INTRODUCTION

There are many kinds of methods to identify artificial ripening tomatoes and natural mature tomatoes. We can see the shape and gently knead tomatoes with hand, to judge by feel (Ma, 2008). The color of natural mature tomato’s appearance is orange red and artificial ripening tomato appearance’s color is bright red; a natural mature tomato around pedicel shows green and artificial ripening tomato seldom (Cheng, 2008). From Fig. 1 we can see, in Fig. 1a, a big red tomato appearance, pedicel around shows khaki and after cutting the left a "hole" present, tell us that the tomato is not fresh and artificial ripening; in Fig. 1b, tomato appearance is big red, pedicel around shows few green and after cutting the right has little juice, the tomato is artificial ripening too; in Fig. 1c, tomato appearance is orange red, pedicel surrounding red and green color, after cutting juicy, red meat, no “hole”, a natural mature tomato. In addition, although some tomatoes’ appearance color and internal structure is normal, the color around pedicel is not normal, this kind of tomatoes is natural mature, not fresh, the consumer will not like it. This study will treat the stale tomato as artificial ripening tomato. Thus it can be seen that, from the tomato outside, appearance color and pedicel surrounding color is main basis to identify artificial ripening tomatoes and natural mature tomatoes. At present in China, mainly use artificial methods to identify artificial ripening tomato. Artificial recognition has not only heavy labor intensity, but also

Corresponding Author: Haibo Zhao, Engineering Technology Research Center of Optoelectronic Appliance, Anhui Province, Tongling Anhui 244000, China

low work efficiency and accuracy. This study adopts automatic identification system with double parallel genetic neural network structure, so as to realize the automatic recognition of artificial ripening tomato and natural mature tomato.

Fig. 1: Comparison of artificial ripening tomato and nature mature tomato
MATERIALS AND METHODS

Hardware of recognition system: Recognition system is as shown in Fig. 2. Computer has IBM compatibles (Intel Celeron 420), 2G memory, 60G hard disk and 128M video memory. Image acquisition card uses CA-CPE-3000 (Zhang et al., 2001; Xie, 2002) from technology group of the Zhongzi, the maximum resolution of the image acquisition display is 768×576×32 bit, image transfer as much as 60 MB/S. The CCD camera uses Panasonic WV-CP480, 1/3 inch CCD, resolution 752 (H)×582 (V), the minimum illumination 0.001Lux, level clear 540 TVL. When testing, around tomato pedicel contact to loophole closely and ensure without gap between tomato and loophole. On both sides of the tomatoes install two mirrors angled 45° with horizontal plane. Then a CCD camera can absorb three sides image information appearance of each tomato and basically guarantee the comprehensive requirements of color detection.

Extraction of color feature: There are a variety of color models when describing color of an object. In the practical application, the commonly used models are RGB and HIS color model (Li et al., 2008; Jau et al., 2008; Zhao et al., 2009; Qin et al., 2008). RGB model, based on the display device, can accurately say color composition on screen. But the component of RGB model has no direct contact with humans' sense of colors; HIS model, based on humans’ mental feeling of color, is in line with people's visual feeling and also the main use of color model in computer vision technology. The RGB model to HIS model conversion formula is:

\[
\begin{align*}
H &= \cos^{-1}\left(\frac{1}{2(R-G)+(R-B)}\right) \\
I &= (R+G+B)/3 \\
S &= 1-3\min(R,G,B)/(R+G+B)
\end{align*}
\]  

where,

R : Red  
G : Green  
B : Blue  
H : Hue  
I : Brightness (intensity)  
S : Saturation

In order to make up the shortages of a single color space representation to color characteristics, this study put a total of 24 variables as artificial ripening tomato quantitative description: the average of tomato external image R, G, B, H, I, S color component (μR, μG, μB, μH, μI, μS) and (μr, μg, μb, μh, μi, μs) and standard deviation (σR, σG, σB, σH, σI, σS) and (σr, σg, σb, σh, σi, σs). Among them, (μR, μG, μB, μH, μI, μS) and (σR, σG, σB, σH, σI, σS) mean tomato appearance color’s feature vector, (μr, μg, μb, μh, μi, μs) and (σr, σg, σb, σh, σi, σs) mean surrounding tomato pedicel color’s feature vector.

Double parallel neural network structure: This study adopts Double Parallel Feed forward Neural Network (DPFNN), Paralleled a single forward Network and a multilayer Feed forward Network. In DPFNN, the output node receives not only the information of hidden unit, but also the information of input layer node directly. So DPFNN is a linear-nonlinear coordinately mathematical model (Guo, 2009; Zhao et al., 2010), as shown in Fig. 3.

Considering the recognition system in this study acquires tomato appearance color and surrounding pedicel color at the same time, neural network uses the parallel hidden layer structure. It groups the collected data, sends the tomato appearance color data to input layer U1 and the data of surrounding pedicel color to input layer U2. Due to the tomato external color feature vector (μR, μG, μB, σR, σG, σB, μH, μI, μS, σH, σI, σS, μr, μg, μb, μh, μi, μs, σr, σg, σb, σh, σi, σs) having different dimension and order of magnitude, to avoid the characteristics of high dynamic range submerging...
the features of low dynamic range, all sample data participating in the analysis will be under normalization (Chen et al., 2009):

\[ c_n = \frac{c - \min(c)}{\max(c) - \min(c)} \]  

(2)

where,  
- \( c \): Tomato appearance color’s feature  
- \( \min(c) \): Minimum value of \( c \)  
- \( \max(c) \): Maximum value of \( c \)  
- \( c_n \): The characteristic value after the normalization

Put new vector \((c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}, c_{12}, c_{13}, c_{14}, c_{15}, c_{16}, c_{17}, c_{18}, c_{19}, c_{20}, c_{21}, c_{22}, c_{23}, c_{24})\) as the input of original input layer in neural network, after data grouping, \((c_{1}, c_{2}, c_{3}, c_{4}, c_{5}, c_{6}, c_{7}, c_{8}, c_{9}, c_{10}, c_{11}, c_{12})\) as the input of input layer U1, \((c_{13}, c_{14}, c_{15}, c_{16}, c_{17}, c_{18}, c_{19}, c_{20}, c_{21}, c_{22}, c_{23}, c_{24})\) as the input of input layer U2. The output of the neural network based on tomato mature property is divided into two kinds of situations: artificial ripening or natural mature. It is expressed as \((0, 1)\) and \((1, 0)\) with the thermometer method. In hidden layer neurons number \(L\) is determined by trial and error method and the process shows in Fig. 4. In the diagram \(W\) is for the weight array from input layer to hidden layer, \(V\) for the weight array from hidden layer to output layer.

**Genetic algorithm:**

**The determination of coding scheme:** In order to overcome the shortcomings of binary code, the real number coding will be used. The real expression can be directly genetically operated on the phenotype of solution, instead of converting the numerical system. This study encode \(W\) and \(V\) as a chromosome string at the same time. The specific coding mode is: first gradate \(W\) and then gradate \(V\).

**The generation of initial population:** Random number generator produces initial population containing \(N\) chromosome string. If the size of the initial population value of \( N \) is too small, it will be easy to fall into local optimal solution; the value of \( N \) is too big, it will reduce the efficiency. When practical application, we can only determine the value of \( N \) based on experience or experimental, generally selected in \((20, 100)\) (Rudolph, 1994).

**The determination of evaluation function:** Evaluation function is defined as:

\[ f_j = \frac{1}{\sum_{i=1}^{s} \sum_{j=1}^{n} (d_{ij} - r_{ij})^2} \]

(3)

where,  
- \( j \): The \( j \)-th evaluation value of chromosome list  
- \( s \): The number of training sample  
- \( n \): The number of neural network output nodes  
- \( d_{si} \): The expected network output of training sample  
- \( r_{si} \): The real network output of training sample

**Selecting operation:** In the fitness proportion method, the selected probability of each individual is proportional to its fitness value, namely:

\[ p_{si} = \frac{f_i}{\sum_{i=1}^{N} f_i} \]

(4)

where,  
- \( p_{si} \): For the selected probability of the \( i \)-th chromosome list  
- \( \sum_{i=1}^{N} f_i \): For the sum value of each individual fitness

**Adaptive crossover and mutation operation:** In this study two-point-cross method is used for cross operation, bitwise variation method for variation operation (Liao et al., 2010). In order to maintain the diversity of population, use adaptive crossover probability \( P_c \) and mutation probability \( P_m \). Computation formula is as follows:

\[
P_c = \begin{cases} 
\frac{P_{c1} - P_{c2}}{f_{\text{max}} - f_{\text{av}}} f' > f_{\text{av}} \\
P_{c1}, \quad f' \leq f_{\text{av}}
\end{cases} \]  

\[
P_m = \begin{cases} 
\frac{P_{m1} - P_{m2}}{f_{\text{max}} - f_{\text{av}}} f' > f_{\text{av}} \\
P_{m1}, \quad f \leq f_{\text{av}}
\end{cases} \]  

(5)  

(6)

where,  
- \( f' \): The fitness value of the bigger one between two intersecting individuals  
- \( f_{\text{av}} \): The average fitness value of each generation group
The characteristic of this algorithm is strong directional sample, is the solution when gradient of error is zero. The objective function is defined as:

\[ G(k) = \frac{1}{\lambda} \sum_{i=1}^{L} \lambda^{i-1} (d(t) - f(t))^2 \]  

(7)

where,

- \( \lambda \) : Weighted forgetting factor and \( 0 \leq \lambda \leq 1 \)
- \( d(t) \) : The desired output of the output node
- \( f(t) \) : The actual output of the output node

The weight update process which using the method of recursive least square as follows:

\[ w_j(k) = w_j(k-1) + g(k)[d(k) - h(k)w_j(k-1)] \]  

(8)

\[ g(k) = \frac{P(k-1)h^T(k)}{\lambda + h(k)P(k-1)h^T(k)} \]  

(9)

\[ P(k) = [P(k-1) - g(k)h(k)P(k-1)]/\lambda \]  

(10)

RESULTS AND DISCUSSION

The process of system test as follows. Collecting tomato’s external image in the test equipment, extracting color characteristic parameters (R, G, B, H, I, S) after image processing, figuring out the average value and standard deviation of color characteristic parameters and normalizing them. After that, cutting tomatoes, judging its mature properties (output values of neural network) from internal quality and then corresponding the data after normalization process with the output values of neural network, as the sample data. At last, training the double parallel feed forward neural network with genetic algorithm and the trained network model can use to forecast the test sample.

Selecting 160 tomatoes, artificial ripening and natural mature every 80, as the training sample. Another selecting 80 tomatoes, artificial ripening and natural mature every 40, as the testing sample. The selection principle of training and testing sample is the sample with enough representative and comprehensive. And the population size takes N = 80 and two hidden neurons number take L = 9 are the training conditions. All sorts of color characteristic value of training and testing samples, as shown in Table 1 and 2.
Identifying different varieties of tomatoes with the trained network and the results shows as Table 3. The test showed that the system whose average correct identification rate was 93.8% identifies better.

Table 4 shows the comparative result of correct recognition rate of identification test to ripening tomato which is comparing the neural network of same structures trained by BP algorithms with trained by genetic algorithms. According to Table 4, the correct recognition rate, the automatic identification of artificial ripening tomato with the neural network trained by genetic algorithms, is obviously higher than the same structure neural network which is trained with BP algorithm.

**CONCLUSION**

Based on the extraction to image information of tomato’s external color, with the average value and standard deviation of the color component of image R, G, B, H, I, S surrounding tomato’s appearance and pedicel as the input of the neural network and the multilayer feed forward neural network adopting genetic algorithm training realized the automatic identification of artificial ripening tomato. Tests showed that the correct recognition rate of the system is 93.8, which laid a certain theoretical and practical basis for further recognition studies of artificial ripening tomato.

**ACKNOWLEDGMENT**

The authors thank the Engineering Technology Research Center of Optoelectronic Appliance, Anhui Province of China for support.

**REFERENCES**


