Regular paper http://dx.doi.org/10.13067/JKIECS.2015.10.2.267

골프공 인식을 위한 OpenCV 기반 신경망 최적화 구조

김강철*

Optimal Structures of a Neural Network Based on OpenCV for a Golf Ball Recognition

Kang-Chul Kim*

요 약

본 논문은 OpenCV 라이브러리를 기반으로 골프공 인식을 위한 신경망의 최적화 구조와 관심영역의 빛의 명도를 계산한다. 개발된 시스템은 전처리, 영상처리, 기계 학습 과정으로 구성되며, 기계 학습과정은 테스트 영상으로부터 골프공과 다른 오브젝트에 대한 Hu의 7 불변 모멘트, 가로 및 세로 비율 또는 면적으로부터 계산된 π를 입력으로 사용하여 다층 퍼셉트론을 기반으로 학습모델을 구한다. 다층 퍼셉트론에 대한 최적의 은닉층과 노드의 수를 결정하도록 모의 실험한 결과 2개의 은닉층과 각 은닉층에 9개의 노드를 가질 때 최대의 인식율과 최소 실행 시간을 얻었다. 그리고 관심영역의 최적 명도는 200으로 계산되었다.

ABSTRACT

In this paper the optimal structure of a neural network based on OpenCV for a golf ball recognition and the intensity of ROI(Region Of Interest) are calculated. The system is composed of preprocess, image processing and machine learning, and a learning model is obtained by multi-layer perceptron using the inputs of 7 Hu's invariant moments, box ration extracted by vertical and horizontal length or π calculated by area of ROI. Simulation results show that optimal numbers of hidden layer and the node of neuron are selected to 2 and 9 respectively considering the recognition rate and running time, and optimal intensity of ROI is selected to 200.

키워드

Neural network, OpenCV, Optimal Structure, Pattern Recognition
신경망, OpenCV, 최적화 구조, 패턴인식

Ⅰ. Introduction

As computer technology has been developing, pattern recognition algorithms provide a reasonable answer for all possible information processing[1]. To easily implement pattern recognition systems in many field, OpenCV which is an open source computer vision library is used[2-4]. The library contains over 500 functions that span many areas. It is mainly used for some advanced image processing, such as feature detection and tracking, motion analysis, object segmentation and recognition and 3D reconstruction.

Neural network[4], [6] is one of the most active fields in the research and application of artificial intelligence and it causes great concern of people

* 전남대학교 전기전자통신컴퓨터공학부(kkc@jnu.ac.krl)
접수일자 : 2014. 12. 03 심사(수정)일자 : 2015. 01. 16 게재확정일자 : 2015. 02. 09

with its unique advantages. Neural network tries to simulate nervous system of human brain from the perspective of bionics, so that machine can have part of function of human brain such as perception, learning, reasoning. An artificial neuron is a computational model inspired in the natural neurons. Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough, the neuron is activated and emits a signal though the axon. This signal might be sent to another synapse, and might activate other neurons as Fig.1 shows.

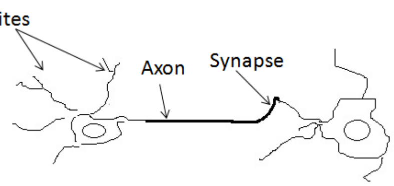


Fig. 1 Natural neurons

Fig.2 shows a neuron model. The complexity of real neurons is highly abstracted when artificial neurons modeled. These basically consist of inputs, which are multiplied by weights, and then computed by a mathematical function which determines the activation of the neuron. Artificial neuron is a simulation of biological neuron, it is a basic processing unit of the neural network. The three elements to constitute BP(back propagation) neuron model are that

(1) A group of synapses, or connections. Commonly, W_{ij} denotes the connection strength between neuron i and neuron j, called weight.

(2) A summation unit Σ, to calculate the weighted sum of input signal.

(3) A excitation function f(*), to limit the output neurons, the output range of the neuron is always [0,1] or [-1,1].

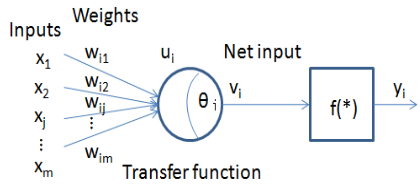


Fig. 2 An neuron model

The above functions can be expressed with the following mathematics

$$u_i = \sum_{j=1}^{M} w_{ij} x_j \tag{1}$$

$$v_i = u_i + \theta_i \tag{2}$$

$$y_i = f(v_i) \tag{3}$$

where x_j(j=1,2,...,m) is the input signal of neuron, w_{ij} is connection weight. u_i is the output of the input signal linear combination, θ_i is the threshold of neurons, v_i is the value with threshold adju-stment. f(*) is BP neural excitation function, the most common form is the sigmoid function like a equation (4).

$$f(*) = (1+e^{-ax})^{-1} \tag{4}$$

The input value of function is between (-∞, ∞), output value is between 0 and 1, and 'a' is constant as shown in Fig.3.

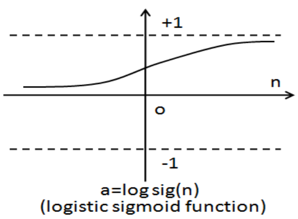


Fig. 3 The excitation function

Hu's moment, box ratio or π from a golf ball image and other objects are used on the training data to the neural network. When using back propagation neural network, the number selection of hidden layer, learning rate and selection of expected error should be determined. If the number of hidden layers is too little, it can't extract the accurate characteristics of samples and can't identify the trained samples, and has a poor fault tolerance. If the number of hidden layers is too large, the structure is complex, and the training time also increases.

The choice of the number of hidden layers and neurons is very important for the effective performance of the pattern recognition. In this paper, to get the optimal structure of a neural network for a golf ball recognition a neural network based on OpenCV is proposed. 7 Hu's moments and box ratio of a golf ball are used as inputs of a neural network. They are obtained by image processing including image capture, ROI extraction, morphology, edge detection, and moment calculation.

This paper is organized as follows. In section II, image processing algorithms are described and the section III describes the proposed algorithms of the neural network by using OpenCV. And The section IV shows the simulation results, finally the conclusion is described in section V.

Ⅱ. Image processing for a golf ball

In this paper, a golf ball recognition system consists in preprocessing, image processing, and golf ball recognition like Fig. 4.

The system inputs the image taken in vision camera for preprocessing. ROI extraction, noise filtering, contour and feature extraction are done for image processing. If these regions can be extracted, then the efficiency and accuracy of processing and analyzing images can be greatly enhanced. Color image is transformed to gray scale image. The

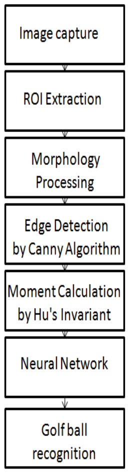


Fig. 4 A recognition system for a golf ball

region which has intensity (the average value of R, G and B) larger than 200 for a golf ball is chosen as the region of interest. Noise included in the result of ROI extraction is filtered and applied to open operation of morphological transformation conducted by the kernel with a random shape and the size that has a fixed point[7]. Four basic operation of Morphology: erosion, dilation, open and close operation[8-9].

To acquire edge information for an input image, Canny edge detection method is applied. After detecting the edge, 7 Hu's invariant moment are used to compare the characteristics of the object. Geometric moments of level (p + q) are defined like a equation (5)

$$m_{pq} = \sum_{i=1}^{n} I(x,y) x^p y^q \quad (p \geq 0, q \geq 0) \tag{5}$$

p and q are the x-order and the y-order, respectively. I(x, y) is the intensity of the pixel (x, y). Hu's moment is the linear combination of unitary central moments with $p+q \leq 3$[10]. Features of normal moment and center moment don't have translation invariance, scaling invariance and rotation invariance. Definitions of the Hu's moments are like equation (6)

$$H_1 = n_{20} + n_{02}$$

$$H_2 = (n_{20} - n_{02})^2 + 4n_{11}^2$$

$$H_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2$$

$$H_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2$$

$$H_5 = (n_{30} - 3n_{12})(n_{30} + n_{12})((n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2 + (3n_{21} - n_{03})(n_{21} + n_{03})$$

$$(3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2)$$

$$H_6 = (n_{20} - n_{02})((n_{30} + n_{12})^2 - (n_{21} + n_{03})^2) + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03})$$

$$H_7 = (3n_{21} - n_{03})(n_{21} + n_{03})(3(n_{30} + n_{21})^2 - (n_{21} + n_{03})^2) - (n_{30} - 3n_{12})(n_{21} + n_{03})$$

$$(3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2) \qquad (6)$$

III. The proposed Neural Network

BP neural network used in OpenCV is a kind of multi-layer feed forward neural network. It is similar to mult-layer perception in structure as shown in Fig.5.

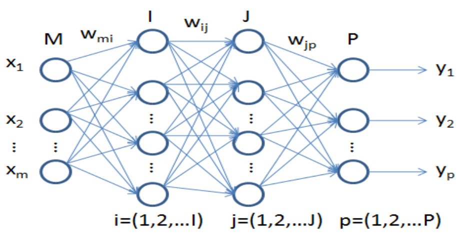


Fig. 5 The structure of neural network

The number of input is m, the number of neurons in the first hidden layer is i, the number of neurons in the second hidden layer is j, the number of output signal is P, the weight between input layer and the first hidden is w_{mi}, the weight between the first hidden layer and the second hidden layer is w_{ij}, the weight between the second hidden layer and output layer is w_{jp}, the input of neuron is x, and the output of neuron is y.

There are many methods to determine the number of hidden layer nodes and it is determined step by step according to the experiment.

Learning rate decides the weight variation generated in circuit training. If the learning rate is set to be too large, it will cause instability of the system. If it is set to be too small, it causes the network to have a slow convergence rate, and the training time will greatly increase. Usually in application, smaller learning rate tends to be selected to ensure the stability of system. The selection of learning rate is usually chosen to be between 0.01 and 0.8.

The purpose of the neural network is to get the learning model which contains the feature of the golf ball. It selects the input and the output data for the supervised learning, and normalizes the input and the output data. And it trains the normalized data using back propagation algorithm and tests the goodness of fit of the model, and finally compares the predicted output with the desired output.

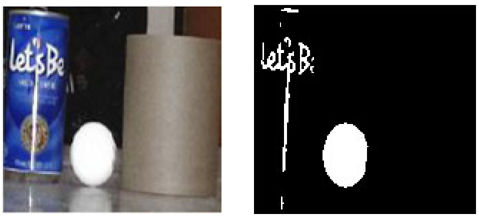
According to results of 7 Hu's invariant moments extracted from an object, the algorithm can recognize whether the object is a sphere like a golf ball or not. The pseudo code is given as Fig. 6.

```
1: Mat frame, frame_split[3], element;
2: frame = imread("test.png"); //read a image form a file
3: cvSplit(frame, frame_split); //split frame into 3 matrix B G R, each matrix has 1 channel
4: if(frame.at<uchar>(i,t) > 200) // ROI extraction (the region with larger intensity than 200 is extracted).
5: frame_split[0].at<uchar>(i,j) = 255;
else frame_split[0].at<uchar>(i,j) = 0;
6: element = cvGetStructuringElement(MORPH_RECT, Size(3,3));
7: cvMorphologyEx( frame_split[0], frame_split[1],MORPH_OPEN, element); //noise filtering
8: cvCanny(frame_split[1], frame_split[2], lowThresh, highThresh, 3) //edge detection
9: cvFindContours( frame_split[2], contours, hierarchy, C
```

```
V_RETR_TREE , CV_CHAIN_APPROX_SIMPLE); //find the object contour
10: double area=cvContourArea(contours[idx]); //contour area calculation
11: if(area > 100) { // ROIs with more than 100 pixels are selected
12.(a): Rect box = cvBoundingRect(contours[idx]); //compute the rectangle boundary
13: Moments mo = cvMoments(contours[idx]); //extract moment
14: double h[7]; //7 hu invariant moment
15: HuMoments(mo, h); //compute the value of Hu's invariant moments
16.(a): float result = Predict(h[0], h[1], h[2], h[3], h[4], h[5], h[6], box_rate); // the output of neural network
17: if(result > 0.9) { //if output is greater than 0.9 ,the system recognize the object as a golf ball
18: cvDrawContours( frame, contours, idx, color, 1, 8, hierarchy); //draw the contour of the object
19: double distance = - pow(area,7)*1.3e-023 + pow(area,6)*4.6e-019 - pow(area,5)*6.2e-015 + pow(area, 4)*4.4e-011 - pow(area,3)*1.7e-007 + pow(area,2)*0.00037 - area*0.43 + 280; //distance calculation
```

Fig. 6 Golf ball recognition algorithm using box ratio

IV. Simulation and Results



(a) Original image (b) ROI extraction

Fig. 7 Original image and ROI extraction

Fig.7 shows an original image (a) and an ROI extracted image (b) that has intensity >200. The original image has three objects including a golf ball and two other objects, a coffee can and a paper can which are seen at in-door golf driving range. Fig.7(b) shows that the region with white or brighter region of more than 200 intensity level has been extracted and chosen.

After Canny edge extraction, the edge of objects is extracted. The system selects the object with the area larger than 100 while finding contours. So the small contours are removed. The result is shown in Fig.8.

Fig. 8 Contour extraction

Table 1. The values of Hu's invariant moments for three objects

	Golf ball	Coffee Can	Paper can
H1	0.159361	0.677755	0.246015
H2	2.84E-05	0.418109	0.0166552
H3	2.49E-06	0.045503	0.00248555
H4	3.16E-09	0.0425293	6.3E-06
H5	7.63E-17	1.87E-03	7.63E-10
H6	8.10E-12	2.75E-05	8.21E-07
H7	2.70E-16	7.17E-06	2.46E-10

Table 1 shows invariant moment values for a golf ball, a coffee can and a paper can. Because of the different shapes, the Hu's invariant moments of the three objects are different. Hu's invariant moment's value decreases as the prior degree increases, and the table shows that the values of these factors are all much smaller than 1.

Table 2. The number of hidden layer and neurons

H	N	T(s)	R
1	5	0.000175	69
	7	0.000186	96
	9	0.000197	131
	11	0.000217	215
	13	0.000239	247
	15	0.000275	184

2	5	0.000353	253
	7	0.000425	306
	9	**0.000536**	**337**
	11	0.000647	202
	13	0.000772	174
	15	0.000963	143
3	5	0.000642	302
	7	0.000772	331
	9	0.000936	335
	11	0.00117	337
	13	0.00134	327
	15	0.00158	296

H : number of hidden layer, N : number of neurons, T : simulation time for each sample, R : number of recognition

Table 2 shows the simulation result of neural network depending on the number of hidden layers and the number of neurons in hidden layer. When the number of hidden layer is 2 and the neuron is 9, the number of recognition for golf ball is the highest and simulation time is short.

Table 3. The determination of the optimal intensity

Intensity=150

Samples \ Result	P	F
Golf ball(60)	57	3
Other Object(60)	44	16

Intensity=200

Samples \ Result	P	F
Golf ball(60)	51	9
Other Object(60)	60	0

Intensity=250

Samples \ Result	P	F
Golf ball(60)	36	24
Other Object(60)	60	0

P=pass, F=fail

To determine the optimal intensity in ROI extraction, 150, 200, and 250 intensities are chosen respectively, and 120 images are used for detecting (60 images for golf balls and 60 for other objects) like Table 3. The accuracy are 84.2%, 92.5% and 80% when the intensities are 150, 200, and 250 respectively. So when the intensity is equal to 200, the number of the golf ball recognition is high.

V. Conclusions

The choice of the number of hidden layers and neurons in neural network is very important to get effective performance of the pattern recognition. In this paper, an optimal structure of a neural network based on OpenCV is proposed for a golf ball recognition. The system is composed of preprocess, image processing and machine learning, and a learning model is obtained by multi-layer perceptron using the inputs of 7 Hu's invariant moments, box ration extracted by vertical and horizontal length or π calculated by area of ROI. Simulation results show that optimal numbers of hidden layer and the node of neuron are selected to 2 and 9 respectively considering the recognition rate and running time, and optimal intensity of ROI is selected to 200 with accuracy of 92.5%.

References

[1] I.-S. Oh, *Pattern Recognition*, Korea, Kyobobook publisher, 2008.

[2] R. Laganière, *OpenCV 2 Computer Vision Application Programming Cookbook*, UK, Packt Publishing, 2011.

[3] G. Bradski, A. Kaebler, *Learning OpenCV: Computer Vision with the OpenCV Library*, CA, O'Reilly Media, 1st edition, October, 2008.

[4] J. Zhu, J. Chong, K-C Kim, "The Recognition

and Distance Estimation of a Golf Ball using a WebCam," *J. of Korea Institute of Electronic Communication Sciences*, vol. 8, no. 12, 2013, pp. 1833-1839.
[5] K.-C. Kim, S.-Y. Lee, S.-Y. Moon, "Design of the recognition and position estimation system for a golf ball collecting robot," *SCTA*, KwangJu, 2012, pp. 104-106.
[6] T. Wanmei, "The Study of the Optimal Structure of BP Neural Network," *Systems Engineering Theory and Practice*, vol. 25, no. 10, 2005, pp. 95-100.
[7] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Prentice Hall, Third edition, 2008.
[8] http://www.cse.iitd.ernet.in/~pkalra/csl783/canny.pdf
[9] X. He, J. Li, D. Wei, W. Jia, Q. Wu, "Canny Edge Detection on a Virtual Hexagonal Image Structure,"*JCPC2009*, Sydney, 2009, pp. 167-172.
[10] M.-K. Hu, "Visual pattern recognition by moment invariants,"*IRE Trans. Information Theory*, Vol. 8, 1962, pp. 179-187.
[11] K.-B Kim, Y.-W Woo, "Content-based Image Retrieval Using HSI Color Space and Neural Networks," *J. of Korea Institute of Electronic Communication Sciences*, vol. 5, no. 2, 2010, pp. 152-157.

저자 소개

김강철(Kang-Chul Kim)

1981년 서강대학교 전자공학과 학사
1983년 서강대학교 전자공학과 석사
1996년 경상대학교 전자공학과 박사
현재 전남대학교 전기전자통신컴퓨터공학부 교수
※ 관심분야 : 임베디드시스템, NoC, Pattern Recognition