Knowledge-Based Numeric Open Caption Recognition for Live Sportscast

Si-Hun Sung
Technical Research Center, Munwha Broadcasting Corp., Seoul 150-728, Korea
shsung@mbc.co.kr

Abstract

Knowledge-based numeric open caption recognition is proposed that can recognize numeric captions generated by character generator (CG) and automatically superimpose a modified caption using the recognized text only when a valid numeric caption appears in the aimed specific region of a live sportscast scene produced by other broadcasting stations. In the proposed method, mesh features are extracted from an enhanced binary image as feature vectors, then a valuable information is recovered from a numeric image by perceiving the character using a multiplayer perceptron (MLP) network. The result is verified using knowledge-based rule set designed for a more stable and reliable output and then the modified information is displayed on a screen by CG.

MLB EyeCaption based on the proposed algorithm has already been used for regular Major League Baseball (MLB) programs broadcast live over a Korean nationwide TV network and has produced a favorable response from Korean viewers.

1. Introduction

Relayed sportscasts produced by foreign television stations include useful game information, such as the game score, speeds, and distances as well as player information for the TV audience. This game information is usually displayed on a screen by making a numeric open caption in units according to the cultural norm.

The international system of units (SI) promoted by the international organization of legal metrology (OIML) has been accepted by most countries as the standard for weights and measures [1]. However, non-SI measurements, such as a mile (mi.), yard (yd.), foot (ft.), and inch (in.), and weights, such an ounce (oz.), pound (lb.), and gallon (gal.), are still more popular in certain countries, including the United States.

When information measured using non-SI units is supplied to viewers who are more familiar with SI units, this creates confusion, and vice versa. Consequently, unfamiliar captions need to be translated into familiar captions according to the audience. Presently, studio operators have to manually recalculate the measures and then display the converted captions on a screen. This results in low efficiency and high fatigue of the operator, plus the conversion speed is slow.

In some broadcast applications related to pattern recognition research, Qi *et al.* [2] employed a video ognition research, Qi *et al.* [2] employed a video optical character recognition (OCR), whose classifier is a linear Support Vector Machines, to assist speech recognition for a content-based news video browser. Natarajan *et al.* [3] developed an image enhancement technique and an OCR engine using hidden Markov models for videotext. We propose a knowledge-based numeric open caption recognition that can perceive a numeric image using an multiplayer perceptron (MLP) network that can produce a stable output for numeric character recognition, and efficiently and rapidly convert the caption only when a valid numeric image appears in a specific region, called *watching region*.

We detail the structure and procedure of the proposed method in Section 2. Section 3 presents real broadcast results of MLB EyeCaption, which is an application implemented using the proposed method. Finally, some conclusions are made in Section 4.

2. Numeric Open Caption Recognition

A block diagram of the proposed numeric caption recognition is shown in Figure 1. First, the size of the character image is normalized by accurately extracting a minimum bounding rectangle (MBR) for each character from the enhanced image in the watching regions, as indicated by the operator in advance. Next, the feature vector extracted from the size-normalized binary character image is input into a MLP network. Finally, the converted caption is generated using character generator (CG), then superimposed on the original scene by video mixing unit (VMU). The knowledge-based rule set verifies the MBR and recognized result.

2.1. Image Enhancement

Usually, a caption is composed of a key output determining the position and transparency of the on-screen graphic character and a video output deciding the color and texture of the character. Translucent and opaque captions are often used together for a more smart view. However, since the background of the translucent caption can have a negative influence on the recognition process, the image needs to be enhanced before feature extraction. In addition, anti-aliasing, a technique which smoothes jagged edges in graphic objects, can have a detrimental effect on the recognition of small characters because it adjusts the brightness of the pixels near the character boundary.

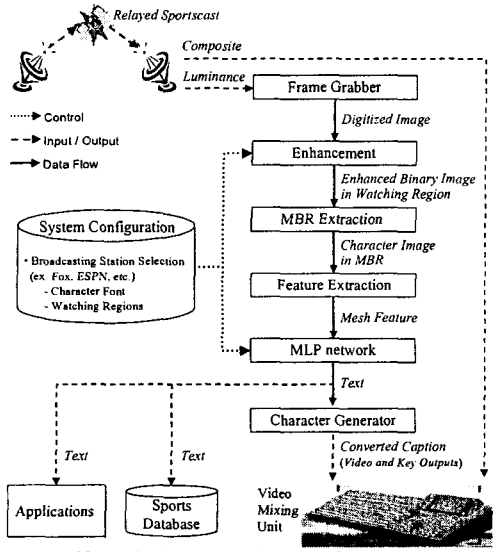


Figure 1. Knowledge-based numeric open caption recognition for live sportscast.

Accordingly, to reduce this influence, the following image processing is used. First, black and white references are established to enhance the contrast in 8-bit A/D. Second, the distinguishable peculiarities among the target characters are enlarged using a morphological operation, which is erosion for a bright character and dilation for a dark one. Finally, a binary image is created to minimize the influence of slight variations in brightness. The threshold T for binarization can be described as

$$T = \max(T_{Otsu}, m), \tag{1}$$

where T_{Otsu} is the threshold according to Otsu's method [4] and m is the mean brightness. When the size of the watching region is much larger than that of the character, there is a strong tendency to select T_{Otsu} for T, and when the sizes are similar to select m for T.

2.2. MBR and Mesh Feature Vector

The MBR of each character, i.e. the bounding rectangle of the largest blob in each watching region, is found out for extracting the more substantial features from the image. A binary labeling method [5] is used so as to accurately segment the character region, then 16×16 size-normalized mesh feature vectors are extracted as the representative feature vectors of the non-overlapped blocks in the binary image to support different character sizes. Since a mesh feature vector represents the mean of each local region divided by a mesh, it includes useful information on the rough shape of the character and effectively decreases the dimensions of the input nodes in the MLP network.

2.3. Neural Network

Neural network, a generalizing methodology that represents a mathematical model of neural biology, consists of simple processing elements called *neurons* or *nodes* that are connected to one another by directed communication links with associated weights [6]. We use a fully connected MLP network, as shown in Figure 2, as the presence of hidden layer is able to solve more complicated problems than a network with only a single layer composed of just input and output units. The MLP network is trained using an error back-propagation (BP) algorithm [7], which is a gradient descent method to minimize the total squared error of the output computed by the network, where the error is estimated by subtracting the actual response of the network from the desired response.

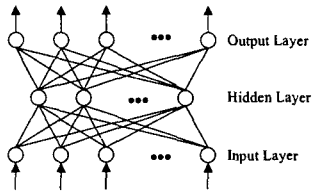


Figure 2. Architecture of MLP network

2.4. Knowledge-Based Rule Set

The knowledge-based rule set designed to create a more stable and reliable output consists of MBR verification rules and recognition verification rules. After extracting a MBR, the MBR verification rules inspect whether the image in the watching region is a valid character image. When the character is brighter than the background, the rules are as follows:

$$\text{Rule 1.} \begin{cases} \text{Reject,} & \text{if } m_{MBR} < m_{min}, \\ \text{Accept,} & \text{otherwise.} \end{cases} \tag{2}$$

$$\text{Rule 2.} \begin{cases} \text{Reject,} & \text{if } W_{MBR} < W_{min}, \\ \text{Accept,} & \text{otherwise.} \end{cases} \tag{3}$$

$$\text{Rule 3.} \begin{cases} \text{Reject,} & \text{if } H_{MBR} < H_{min}, \\ \text{Accept,} & \text{otherwise.} \end{cases} \tag{4}$$

$$\text{Rule 4.} \begin{cases} \text{Reject,} & \text{if } \frac{W_{MBR}}{H_{MBR}} > \gamma_{max}, \\ \text{Accept,} & \text{otherwise.} \end{cases} \tag{5}$$

$$\text{Rule 5.} \begin{cases} \text{Reject,} & \text{if } W_{MBR} H_{MBR} = \alpha, \\ \text{Accept,} & \text{otherwise.} \end{cases} \tag{6}$$

Rule 1 tests whether the character image in the MBR is sufficiently bright to identify, where m_{MBR} is the mean brightness in the MBR and m_{min} is the allowed minimum value. Rules 2 and 3 reject the recognition of the image with a tiny MBR, where W and H denote the width and height of the MBR, respectively, In addition, the aspect ratio of the MBR is verified by Rule 4, whereby only an aspect ratio less than the maximum aspect ratio γ_{max} is allowed. In Rule 5, when the estimated MBR is exactly the same size as the watching region α, the whole image in the watching region is regarded as background and thus rejected.

After recognizing a numeric character, the recognition verification rules are applied to certify the validation of the recognized output as follows:

$$\text{Rule 6.} \begin{cases} \text{Accept,} & \text{if } \dfrac{O_{first}}{O_{second}} \geq \sigma_{min}, \\ \text{Reject,} & \text{otherwise.} \end{cases} \quad (7)$$

$$\text{Rule 7.} \begin{cases} \text{Accept,} & \text{if } n_{min} \leq n \leq n_{max}, \\ \text{Reject,} & \text{otherwise.} \end{cases} \quad (8)$$

The reliability factor, defined as O_{first}/O_{second} for the winner of the output neurons, has to be larger than the minimum reliability factor σ_{min} for the recognition to be confirmed, as shown in Rule 6, where O_{first} and O_{second} represent the output of the winner node and second winner node, respectively, in the output layer of the network. When the numeric character of interest is numbers of more than two digits, the final output n is made up of the results of each watching region and must be within a valid range $[n_{min}, n_{max}]$, as evident from Rule 7.

3. Experimental Results

MLB EyeCaption, which is able to automatically convert captions in miles per hour (MPH; mi./h), as used for the pitching speed of a pitcher in the United States, into captions in kilometers per hour (KPH; km/h), which is more familiar for Koreans, was developed using the pro posed algorithm and applied to regular Major League Baseball (MLB) live broadcasts relayed from broadcasting stations in the United States.

Figure 3 shows some examples that include quite different situations, including a scene with a normal graphic scoreboard in Figure 3(a), taken from the MLB programs. As shown in Figure 3, the scenes also include unexpected situations, such as the spectators and baseball ground appearing in the watching region after the pitching speed caption vanishes, distorted scenes due to the fast motion of the TV camera, interlaced TV scanning, fade effect, and other 3D effects. Moreover, in Figure 4, the speed captions themselves are also distorted on account of interlaced scanning and the fade effect, and can share the region coming into view with other information, for example, the ball count and out count.

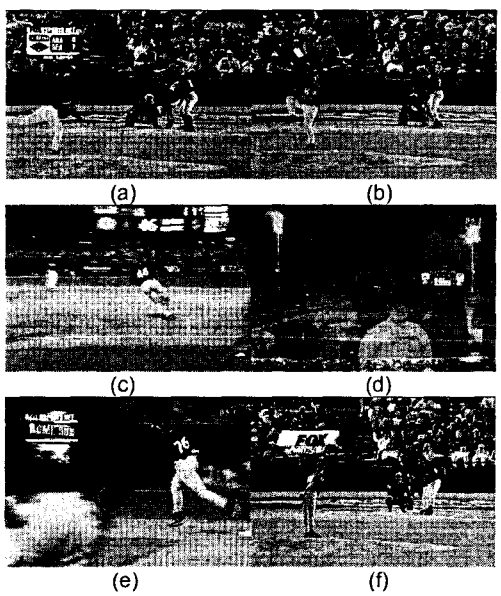


Figure 3. Example scenes from MLB broadcast: (a) normal scoreboard, (b) background composed of spectators, and distorted scenes due to (c) fast motion of camera, (d) fade effect, (e) effect near watching region, and (f) 3D effect of scoreboard.

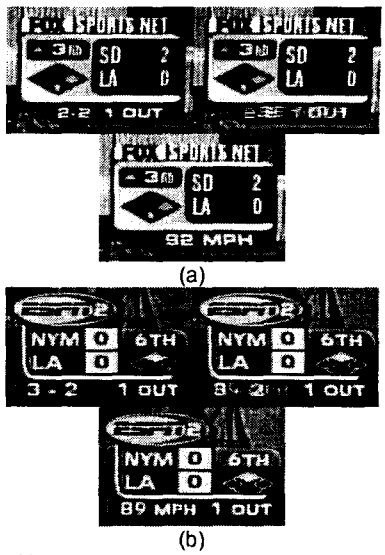


Figure 4. Transient image distortion due to (a) interlaced TV scanning and (b) fade effect.

Figure 5 shows examples of the enhanced binary image and mesh feature for the graphic characters used by Fox Sports Net, one of the MLB broadcasting stations. Characters of Figure 5(a) smaller than 12×12 pixels were treated by the above-mentioned anti-aliasing. In addition, Tables 1 and 2 describe the number of training data for Fox Sports Net and ESPN, respectively.

In the experiments, our classifier was able to identify Arabian numerals from 0 to 9 and the MLP network was composed of 256 sensory nodes in the input layer, 25

Figure 5. Examples of (a) original images in MBR, (b) enhanced binary images, and (c) mesh features.

Table 1. Number of training data for Fox Sports Net.

Target	0	1	2	3	4	5	6	7	8	9	Total
# Data	194	221	218	212	170	184	246	220	257	277	2199

Table 2. Number of training data for ESPN.

Target	0	1	2	3	4	5	6	7	8	9	Total
# Data	340	343	337	332	330	355	369	338	327	329	3400

neurons in the hidden layer, and 10 neurons in the output layer. The number of hidden neurons was determined by trial and error to achieve the best performance.

The proposed classifier produced recognition errors in 56 frames out of 317599 frames up to the ninth inning; therefore the error rate was 0.018%. Similar results were also achieved from tests on other games produced by other broadcasting stations. In one frame, the scoreboard disappeared from the screen and the system mistook the background as characters. In the other 55 frames, the classifier mistook '5' for '8' because of the fading transition to other captions. However, these transient errors could be removed by the vote of five recent recognized outputs.

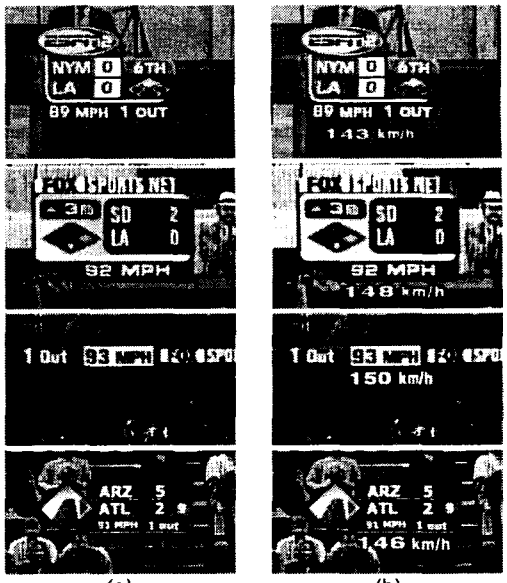
MLB EyeCaption was able to accurately convert the caption in MPH into in KPH only when the pitching speed was appeared (274 times) throughout the entire game up to the ninth inning. It was assumed that the pitching speed was two digit numbers, as such n_{min} and n_{max} in Rule 7 were set at 50 and 99, respectively.

Figure 6 shows the final result in KPH superimposed on the original scene. In fact, the proposed MLB EyeCaption has already been used for regular live MLB programs broadcast over a Korean nationwide network with a good response from viewers.

4. Conclusions

We presented a knowledge-based numeric open caption recognition that can perceive the unfamiliar measures in live sportscasts produced by foreign broadcasting stations and convert these captions into more familiar measurements, plus MLB EyeCaption was developed for live MLB broadcasts in Korea. Mesh features are extracted from the enhanced binary image, then a MLP network is used as the pattern classifier. Furthermore, by defining simple knowledge-based rule set, the proposed method can effectively distinguish only numeric images of interest from a wide variety of typical images and refine the final recognition results. A sports database could also be built using the recognized data to supply viewers with more extensive game information.

Recently, there has been a lot of computer vision research related to pattern recognition and augmented reality to overcome the limits of computer graphics techniques in broadcasting systems. As such, MLB EyeCaption, which is the first on-line equipment using the character recognition for regular live sportscasts, is an interesting accomplishment in broadcast.



(a) (b)

Figure 6. Experimental results: (a) original scenes and (b) broadcast scenes with converted caption in KPH superimposed on (a).

References

[1] SI 7th Ed. Editorial Committee, *The International System of Units*, Korea Research Institute of Standards and Science, 1998.

[2] W. Qi, L. Gu, H. Jiang, X. Chen, and H. Zhang, "Integrating Visual, Audio and Text Analysis for News Video," *Proc. IEEE International Conference on Image Processing*, Vol. 3, pp. 520-523, 2000.

[3] P. Natarajan, B. Elmieh, R. Schwarz, and J. Makhoul, "Videotext OCR Using Hidden Markov Models," *Proc. The Sixth International Conference on Document Analysis and Recognition*, pp. 947-951, 2001.

[4] N. Otsu, "A Thresholding Selection Method from Gray-Level Histogram," *Pattern Recognition*, Vol.19, pp.41-47, 1986.

[5] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, 1989.

[6] J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing, 1992.

[7] S. E. Fahlman, "Fast Learning Variations on Back-propagation: An Empirical Study," *Proc. Connectionist Models Summer School*, pp.38-51, 1988.