

스포츠 중계 화면 내 숫자영역에 대한 실시간 문자인식 시스템 구현

성 시 훈, 전 우 성
문화방송 기술연구소

Implement of Realtime Character Recognition System for Numeric Region of Sportscast

Si-Hun Sung and Woo-Sung Chun
Technical Research Center, Munhwa Broadcasting Corporation
shsung@mbc.co.kr

Abstract

We propose a realtime numeric caption recognition algorithm that automatically recognizes the numeric caption generated by computer graphics (CG) and displays the modified caption using the recognized resource only when a valuable numeric caption appears in the aimed specific region of the live sportscast scene produced by other broadcasting stations. We extract the mesh feature from the enhanced binary image as a feature vector after acquiring the sports broadcast scenes using a frame grabber in realtime and then recover the valuable resource from just a numeric image by perceiving the character using the neural network. Finally, the result is verified by the knowledge-based rule set designed for more stable and reliable output and is displayed on a screen as the converted CG caption serving our purpose.

At present, we have actually provided the realtime automatic mile-to-kilometer caption conversion system taking up our algorithm for the regular Major League Baseball (MLB) program being broadcasted live throughout Korea over our nationwide network. This caption conversion system is able to automatically convert the caption in mile universally used in the United States into that in kilometer in realtime, which is familiar to almost Koreans, and makes us get a favorable criticism from the TV audience.

I. 서론

외국 방송사에서 제작한 월드컵축구경기, 국제올림픽경기, 미국 메이저리그 야구경기(Major League Baseball; MLB) 등의 주요 해외 스포츠 중계 영상에는 선수 정보를 비롯한 점수, 속도, 거리 등의 유용한 경기 정보들이 그래픽 자막으로 처리되어 있다. 우리나라는 1964년부터 국제법정계량기구(international organization of legal metrology; OIML)가 권고한 국제단위계(international system of units; SI)를 법정계량단위로 공식적으로 사용하고 있다 [1]. 하지만 미국 등의 국가는 자국 내에서 길이는 마일(mile; mi.), 야드(yard; yd.), 피트(feet; ft.), 인치(inch; in.), 질량은 온스(ounce; oz.), 파운드(pound; lb.), 부피는 갤런(gallon; gal.) 등의 국제단위계가 아닌 단위가 생활에 익숙해져 있어서 현재도 이 단위들을 주로 사용한다. 국제단위계를 사용하지 않은 중계화면을 여과없

이 국내에 방송하였을 때, 자국 시청자들은 문화적 이질감으로 인해 직관적으로 그 정도를 가늠하기가 어렵다. 아울러, 생산활동, 상거래 등 모든 분야에서 법정 계량 단위만을 사용할 것을 권고하는 <계량에 관한 법률> 제3조, <국가표준기본법> 제2조 등의 정부시책에도 부합하지 않는다.

화면 내의 자막을 자국민에게 익숙한 단위 또는 한글 자막으로 변환하기 위해서는 해당 자막이 표출될 때마다 스포츠 중계센터 내의 작업자가 직접 개입해서 작업해야 한다. 하지만 지속적이고 반복적인 작업으로 인해서 작업 효율이 저하되고 작업자의 피로도가 높아질 뿐만 아니라 자막을 실시간으로 변환할 수 없다.

문자인식기술은 근래에 자동차번호판 인식 또는 공장 자동화를 위한 부품인식 등에 활용되고 있으며, 이들 내부 인식기는 주로 템플릿 매칭(template matching)이나 신경회로망으로 구성된다 [2-5]. 본 논문은 숫자에 대해 인식률이 안정되어 있는 신경회로망을 기반으로 스포츠 중계방송에서 보다 효율적이고 신속히 영상의 자막을 변환할 수 있는 실시간 광학 문자인식 알고리즘을 제안한다. 이 알고리즘은 스포츠 중계 화면의 특정 영역에 유효한 숫자 영상이 나타났을 경우에만 작업자의 개입 없이 문자정보를 실시간으로 인식한 후에 사용자의 요구에 맞게 가공해서 자막으로 화면에 출력한다. 논문의 구성은 2장에서 실시간 광학 문자인식 알고리즘의 개요와 세부 블록에 대해서 서술하며, 3장에서 본 논문의 알고리즘으로 구현한 실시간 마일-킬로미터 투구 구속 자동변환 시스템을 실험한 결과를 보인다. 마지막으로 4장에서 논문의 결론을 맺는다.

II. 실시간 광학 문자인식

실시간 광학 문자인식 알고리즘의 전체적인 블록도는 그림 1과 같다. 외국 방송사에서 제작한 스포츠 중계 화면을 위성이나 광케이블로 수신해서 영상획득장치(frame grabber)에서 256레벨의 휘도영상으로 A/D 변환을 한다. 사전에 설정한 인식영역(watching region) 내의 영상을 전처리한 후 최소의접사각형(minimum bounding rectangle; MBR)을 구함으로써 정확한 문자 영역을 찾고 개별 문자영상의 크기를 정규화한다. 정규화 이진 문자 영상으로부터 특정벡터를 추출해서 신경회로망(neural

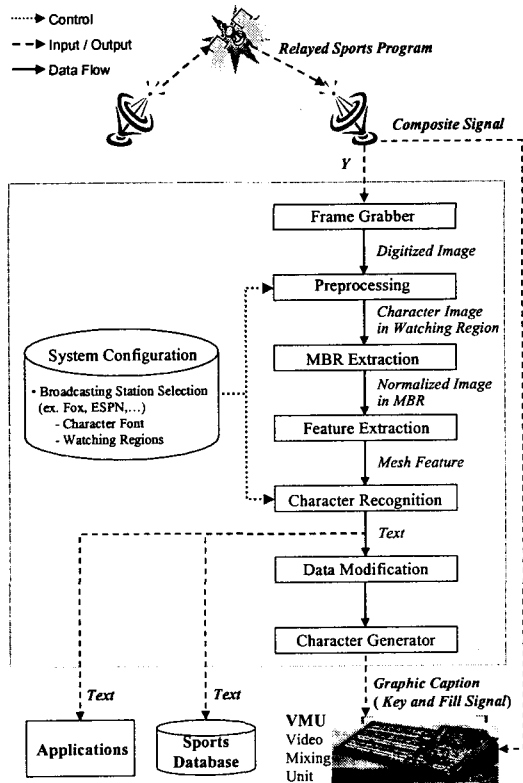


그림 1. 스포츠 중계용 실시간 문자인식 시스템

network)에 입력한다. 인식 결과는 해당 프로그램의 변환자막 뿐만 아니라 다른 응용 시스템, 스포츠 데이터베이스의 자료로 활용된다. 선형적 지식에 기반한 규칙(knowledge-based rule)으로 각 부분에서 검증을 거쳐 변환된 정보는 최종적으로 문자발생기에서 컴퓨터 그래픽(computer graphics; CG)으로 제작되어 영상합성기(video mixing unit; VMU)를 통해 원영상과 함께 표출된다.

2.1 영상 전처리

TV 화면의 자막은 화면에서 자막의 위치와 배경에 대한 자막의 투명도를 결정하는 키(key) 신호와 자막의 RGB 색상을 결정하는 채움(fill) 신호로 구성된다. 일반적으로 자막은 화면에 불투명하게 처리되지만 때로는 아래의 배경과 자막을 반투명도로 표현하기도 한다. 이러한 배경 영상들을 적절히 처리하지 않는다면 자막 인식 과정에서 부적절한 오류를 일으킬 수 있다. 아울러, 자막을 제작할 때 일반화된 컴퓨터 그래픽 기법인 안티알리어스(anti-alias)를 적용한다. 안티알리어스는 그래픽 이미지의 외곽선에 중간색조를 부가해서 자막 가장자리의 계단현상을 제거하고 인공적인 이미지를 부드럽게 한다. 이 안티알리어스는 문자주변의 밝기값을 수정하기 때문에 문자가 비교적 큰 자막의 인식에서는 영향이 크지 않지만 상대적으로 자막의 크기가 작을 경우는 인식에 영향을 미칠 수 있다.

따라서 영상을 획득한 후 앞서 기술한 인식 저하 요인을 제거하기 위해서 다음의 영상처리를 수행한다. 첫째, 하드웨어적으로 영상을 8비트 디지털 영상으로 변

환하는 과정에서 하위밝기 기준레벨(black reference)과 상위밝기 기준레벨(white reference)을 설정함으로써 대조비(contrast)가 높아지도록 히스토그램(histogram)을 조작한 효과를 구현한다. 둘째, 작은 문자 자막은 2, 5, 6, 9의 끊어진 부분의 간격이 2화소 정도로 작기 때문에 수축(erosion) 처리를 해서 문자간 특징벡터의 변별력을 높여준다. 마지막으로 문자의 특징을 잘 표현할 수 있도록 개별 MBR을 구하고 프레임에 따라 변하는 미소한 밝기레벨의 영향을 줄이기 위해서 이진화를 수행한다.

2.2 MBR과 특징벡터 추출

사용자가 지정한 영역에서 정확한 문자 영역을 분할(segmentation)하는 방법으로 이진 레이블링(labeling)[6]을 주로 사용한다. 이진화를 위한 문턱값(threshold) T 는

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

으로 구한다. 여기서, T_{Otsu} 는 Otsu가 제안한 방법 [7]으로 구한 문턱값이고, m 은 평균밝기값이다. 글자 크기에 비해서 인식 영역이 비교적 크게 설정되었을 경우에 T_{Otsu} 가 문턱값으로 채택되고, 크기가 비슷한 경우는 m 이 선택되는 경향을 보였다. 분할된 영역으로부터 문자 영역의 MBR을 구하고 MBR 내부의 영상을 정규화한다. 크기 정규화된 영상으로부터 특징벡터로 3x3화소의 그물망 특징(mesh feature)을 추출한다. 그물망 특징은 그물망처럼 국부 영역으로 나뉘어진 각 영역의 평균값들을 대표벡터로 사용함으로써 문자 모양에 대한 정보를 획득할 뿐만 아니라 입력벡터의 차원을 줄일 수 있다.

2.3 신경회로망

신경회로망은 컴퓨터를 이용해서 생물체 두뇌신경조직의 기본 단위인 뉴런(neuron)을 포함한 신경체계를 수학적으로 모델링하는 기술이다. 뉴런은 노드(node), 셀(cell), 유닛(unit)이라고도 한다. 신경회로망 모델로 하나의 은닉층을 가진 다층 퍼셉트론(multi-layer perceptron; MLP)을 사용하였으며 그 구조는 그림 2와 같다. 그리고, MLP 내 뉴런들간의 연결 강도, 즉 가중치(weight)를 조정하는 알고리즘은 오류 역전파(backpropagation; BP) 알고리즘 [8,9]을 사용하였다. BP 알고리즘은 신경회로망에서 출력층의 오차가 최소화되도록 경사 추적 방법을 사용하여 가중치를 조정하는 방식으로서, 출력층에서 바로 아래의 은닉층으로 오차를 역전파해서 가중치를 조정하고 은닉층에서는 연결된 상위층으로부터 오차를 추정된 후 다시 입력층으로 오차를 역전파해서 입력층과 연결된 가중치를 조정한다.

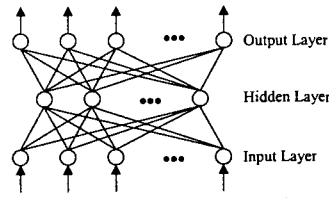


그림 2. MLP 신경회로망

2.4 선형적 지식에 기반한 규칙

선형적 지식에 기반한 규칙은 MBR 검증 규칙(MBR verification rule)과 인식 검증 규칙(recognition verification

rule)으로 구분한다. MBR 검증 규칙은 문자인식을 수행하기 전에 인식 설정 영역 내의 영상이 올바른 숫자 영상인지를 검증하며 어두운 배경에 밝은 문자일 경우에 아래와 같이 규칙 1에서 규칙 5로 정의한다.

$$\text{규칙 1. } \begin{cases} \text{Reject} & \text{if } m_{MBR} < m_{min} \\ \text{Accept} & \text{otherwise} \end{cases} \quad (2)$$

$$\text{규칙 2. } \begin{cases} \text{Reject} & \text{if } W_{MBR} < W_{min} \\ \text{Accept} & \text{otherwise} \end{cases} \quad (3)$$

$$\text{규칙 3. } \begin{cases} \text{Reject} & \text{if } H_{MBR} < H_{min} \\ \text{Accept} & \text{otherwise} \end{cases} \quad (4)$$

$$\text{규칙 4. } \begin{cases} \text{Reject} & \text{if } \frac{W_{MBR}}{H_{MBR}} > \gamma_{max} \\ \text{Accept} & \text{otherwise} \end{cases} \quad (5)$$

$$\text{규칙 5. } \begin{cases} \text{Reject} & \text{if } W_{MBR}H_{MBR} = \alpha \\ \text{Accept} & \text{otherwise} \end{cases} \quad (6)$$

규칙 1은 MBR 영역 내 화소 밝기의 건전성을 검증하는 규칙으로서 m_{MBR} 은 MBR 내의 평균밝기값이고, m_{min} 은 허용하는 최소 평균밝기값이다. 규칙 2와 규칙 3은 과도하게 작은 MBR을 제거하는 규칙으로서 W 는 가로크기, H 는 세로크기이다. 규칙 4는 결정된 MBR이 적절한 가로대 세로비(aspect ratio)를 가지고 있는지를 검증하는 것으로서 γ_{max} 은 최대 허용비율이다. 규칙 5는 인식영역보다 큰 배경영상을 제거하는 규칙으로 α 는 MBR을 구하기 전의 설정영역 넓이이다.

인식 검증 규칙은 신경회로망으로 인식을 수행한 후에 구한 인식 결과의 건전성을 검증하는 규칙으로 규칙 6과 규칙 7이 해당된다.

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

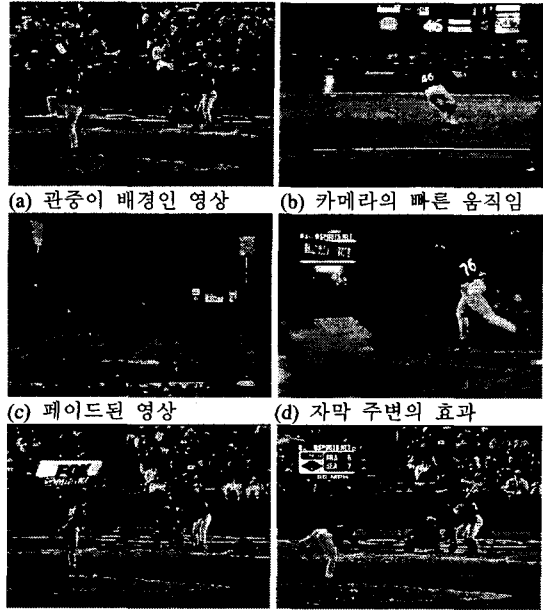
$$\text{규칙 7. } \begin{cases} \text{Accept} & \text{if } n_{min} \leq n \leq n_{max} \\ \text{Reject} & \text{otherwise} \end{cases} \quad (8)$$

규칙 6은 신뢰도(reliability factor)를 평가하는 규칙으로 최소 신뢰값 σ_{min} 보다 커야 한다. 여기서, O_{first} 와 O_{second} 는 각각 출력 노드들 중의 최대 출력값과 두번째 높은 출력값이다. 여러 영역에서 인식된 문자를 조합해서 구해진 결과 n 이 규칙 7에 의해 사전에 설정된 범위 $[n_{min}, n_{max}]$ 를 만족할 때만 결과로 허용한다.

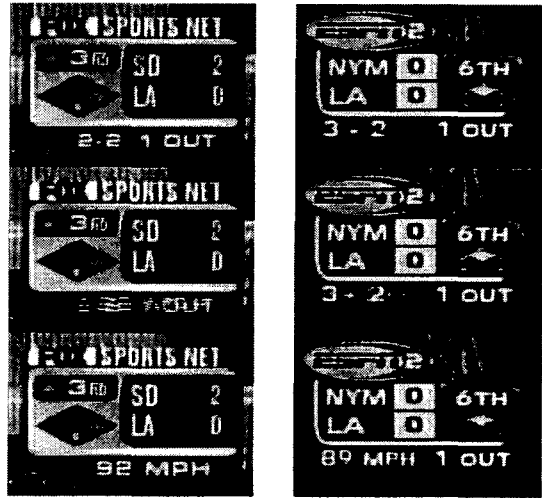
III. 실험 및 결과

상기 설명한 알고리즘을 기반으로 실시간 마일.킬로미터 투구 구속 자동변환 시스템을 구현하였고 실황 위성영상 및 녹화 테이프 등의 MLB 중계영상을 이용해서 경기 시작부터 종료까지 실험하였다.

그림 3은 스포츠 중계에서 볼 수 있는 다양한 영상들 중 일부이다. 그림 3(f)의 정상 영상을 비롯해서 인식 대상 문자인 구속이 표시되는 점수판 자막이 사라지고



(a) 관중이 배경인 영상 (b) 카메라의 빠른 움직임
(c) 페이드된 영상 (d) 자막 주변의 효과
(e) 자막의 3D 돌출 효과 (f) 정상 자막
그림 3. MLB 중계 영상의 예



(a) 주사방식에 의한 왜곡 (b) 페이드에 의한 왜곡
그림 4. 점수판 자막 전환 상태의 영상 왜곡 현상

관중 또는 운동장이 화면에 노출되거나 카메라의 팬(pan), 틸트(tilt), 줌(zoom)과 페이드(fade) 효과, TV의 비월주사방식(interlaced scanning)으로 인해서 생기는 영상의 왜곡, 영상 내 3D 효과, 광고 등의 각종 예측할 수 없는 상황이 발생한다. 더구나 그림 4와 같이 자막 자체도 비월주사방식과 페이드 효과에 의해 왜곡되며, 일반적으로 투구 구속이 표시되는 위치는 볼카운트와 아웃카운트 같은 자막과 영역을 공유하고 있다.

인식기에서 인식할 수 있는 패턴의 종류는 아라비아 숫자 10종이고 MLP 신경회로망 시스템은 256개 노드의 입력층, 25개 노드의 은닉층, 10개 노드의 출력층으로 구

성되었다. 은닉층 노드의 개수는 규모가 비교적 작으면 서도 우수한 결과를 보이도록 실험적으로 결정된 것이다. 그림 5는 인식 대상 문자 중 폭스 스포츠 네트(Fox Sports Net)의 그래픽 폰트에 대한 정규화 이진 영상과 그물망 특징의 예이다. 그림 5(a)와 같이 12x12화소로 작고 안티알리어스로 처리되어 이진화가 쉽지 않다. 표 1은 폭스 스포츠 네트의 해당 문자에 대한 신경망 학습 데이터의 개수이고 표 2는 ESPN에 대한 것이다.

9회까지 수행한 실험에서 총 317599프레임 중 56프레임에서 오인식을 일으켜서 0.018%의 오인식률을 보였다. 1개 프레임에서 점수판이 없어져서 인식 영역 내에 배경만 있음에도 문자로 판별해서 오인식하였고, 자막을 전환할 때의 페이드 효과 때문에 55개 프레임에서 8을 5로 오인식하였다. 이 오류는 각 프레임 출력 후에 후처리를 통해서 제거한다. 전체 경기에서 투구속도가 표시된 횟수는 274번이었고 구속 자막이 나타날 때만 정확하게 킬로미터로 환산해서 자막으로 출력하였다. 다른 여러 차례의 경기 중계에서도 유사한 결과를 보였다.

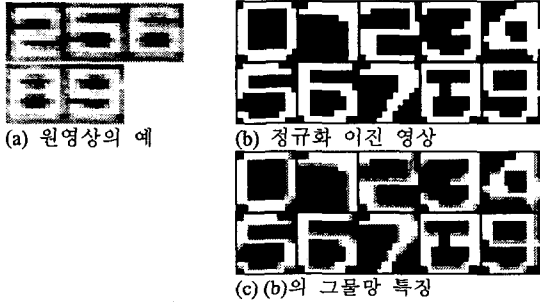


그림 5. MBR 내의 정규화 이진 영상과 그물망 특징

표 1. 폭스 스포츠 네트에 대한 학습 데이터 개수

인식대상	0	1	2	3	4	
데이터 수	194	221	218	212	170	
인식대상	5	6	7	8	9	합계
데이터 수	184	246	220	257	277	2199

표 2. ESPN에 대한 학습 데이터 개수

인식대상	0	1	2	3	4	
데이터 수	340	343	337	332	330	
인식대상	5	6	7	8	9	합계
데이터 수	355	369	338	327	329	3400

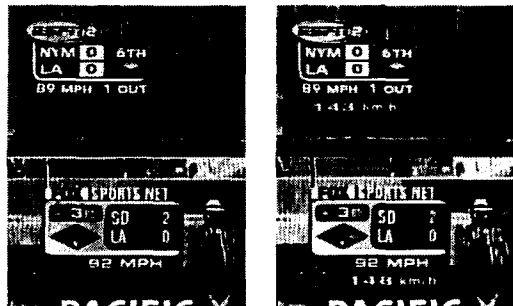


그림 6. 입력 영상의 점수판 영역과 자막 출력

투구 속도의 범위를 50마일에서 99마일까지로 한정하고 인식 대상 영역을 십단위와 일단위로 분리하였다. 최종 자막의 정확도를 높이기 위해서 지난 프레임의 출력값으로 후처리를 한 후에 그림 6(b)와 같이 변환된 값을 자막으로 나타낸다. 이 자동변환 시스템으로 현지 MLB 중계에서 마일로 표시한 구속을 킬로미터로 실시간 변환해서 현재 국내 전역에 생방송하고 있으며, 시청자에게 호응을 얻고 있다.

IV. 결론

스포츠 중계화면에서 우리 환경에 적합한 CG 자막으로 실시간 자동변환하는 알고리즘을 제안하였다. 영상 획득장치에서 얻어진 휘도영상으로부터 MBR 내의 그물망 특징을 추출하였고, 패턴인식기로 MLP 신경회로를 사용하였다. 이 때, 선형적 지식에 기반한 규칙을 정의하여 숫자가 포함된 영상과 그렇지 않은 영상을 용이하게 판별하고 최종 출력의 건전성을 검증하였다. 아울러 시청자에게 보다 풍부한 정보를 제공하기 위해서 변환된 구속 정보는 스포츠 데이터베이스에 저장된다.

호스트(host) 시스템의 하드웨어적인 발전으로 미국, 이스라엘, 프랑스 등에서 군사적 목적으로 주로 연구되던 영상처리와 패턴인식에 관련된 컴퓨터비전 관련 연구들을 방송 시스템에 응용하는 기술이 적극적으로 활성화되는 시점에서 최초로 정확성과 신속성을 요구하는 정규 스포츠 생방송에 문자인식 자막시스템을 고정적으로 활용하고 있으며, 보다 넓은 방송 영역에 컴퓨터비전 관련 기술을 적용할 수 있는 가능성을 보였다.

참고문헌

1. 국제단위계 제7개정판 편집위원회, 국제단위계, 한국표준과학연구원, 1998.
2. E. R. Lee, P. K. Kim, and H. J. Kim, "Automatic Recognition of a Car License Plate Using Imate Processing," *Proc. IEEE International Conference Image Processing*, Vol. 2, pp. 301-305, 1994.
3. P. Comelli, P. Ferragina, M. N. Granieri, and F. Stabile, "Optical Recognition of Motor Vehicle License Plates," *IEEE Trans. Vehicular Technology*, Vol. 44, pp. 790-799, 1995.
4. J. Barroso, E. L. Dagless, A. Rafael, and J. Bulas-Cruz, "Number Plate Reading Using Computer Vision," *Proc. IEEE International Symposium on Industrial Electronics*, Vol. 3, pp. 761-766, 1997.
5. 조보호, 정성환, "ART2 신경회로망을 이용한 차량번호판 문자인식," *한국정보과학회 가을 학술발표논문집*, 제 24권 2호, pp. 455-458, 1997.
6. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, 1989.
7. N. Otsu, "A Thresholding Selection Method from Gray-Level Histogram," *Pattern Recognition*, Vol. 19, pp.41-47, 1986.
8. J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing, 1992.
9. S. E. Fahlman, "Fast Learning Variations on Back-propagation: An Empirical Study," *Proc. Connectionist Models Summer School*, pp.38-51, 1988.