

MPEG Stream Data에서의 온라인 문자인식

이진숙[○] 장춘서
금오공과대학교 컴퓨터공학과
{islee, csjang}@cespc1.kumoh.ac.kr

On-line Character Recognition from MPEG Stream Data

Jin-Suk Lee[○] Choon-Seo Jang
Dept. of Computer Engineering, Kumoh National University of Technology

요 약

본 논문에서는 Web 기반의 원격 교육 환경에서 강사와 학습자 모두에게 도움을 줄 수 있는 판서장면 MPEG Stream Data에서의 온라인 문자 인식 방법에 대하여 연구하였다. 강사가 별도의 프리젠테이션 자료를 만들 필요 없이 직접 판서한 MPEG Stream Data로부터 초당 3 Frame을 Sampling 한 후, 각 Frame에 Laplacian 마스크를 이용한 윤곽선 검출, Frame간 빼기 그리고 세션화 등의 영상처리 기법을 적용하여 문자인식에 필요한 좌표 값과 방향코드 등의 특징을 추출하였다. 좌표 값은 세션화 된 획의 중간 Pixel의 좌표 값이며, 구해진 좌표 값들을 이용하여 8방향 코드와 가상 획 코드를 구한 다음, 이 특징들을 사용해 은닉 마르코프 모델(Hidden Markov Model)을 학습시키고 한글 문자 인식을 행하였다.

1. 서론

최근 인터넷 환경의 비약적인 발전으로 대학을 비롯한 기타 교육 기관에서의 Web 기반의 원격 교육의 수가 급속히 증가하고 있다. 이러한 원격 교육의 강의 방법은 강사, 강의 환경 그리고 강의 내용에 따라 달라 질 수 있는데, 전자 칠판과 프리젠테이션 도구 및 이들의 동기화 기능과 방송 기능을 포함하는 원격 교육 시스템의 경우, 교육 효과를 높일 수 있으나 고가이며 강사가 프리젠테이션 자료를 미리 준비 해야한다. 반면에 강사가 일반 칠판에 직접 판서하는 내용을 음성과 함께 Stream Data로 전송하고, 학습자가 PC에서 판서 내용을 온라인 문자 인식하는 도구를 실행시키며 그 내용을 본다면, 강사와 학습자 모두에게 도움을 줄 수 있을 것이다.

현재의 온라인 인식(온라인 필기 인식)에 대한 입력 방법의 주 연구 분야는 태블릿, 무선(cordless)전자 펜 그리고 액정 화면(liquid crystal display)등에 치중되어 있으며[1], 동영상을 입력으로 사용한 온라인 문자 인식 분야는 드문 실정이다. 그러므로, Web 기반의 멀티미디어 환경을 바탕으로 최근 급증하는 원격 교육 시의 판서 장면을 입력으로 사용한 문자 인식 방법에 관한 연구가 필요하다 할 수 있겠다.

본 논문에서는 원격 교육 환경에서 복잡한 시스템 구성을 요하지 않고 강사와 학습자에게 도움을 줄 수 있는 강의 장면 동영상에서 판서하는 부분에 대한 문자 인식에 관하여 연구하였다.

논문의 구성은 다음과 같다. 2장에서는 본 연구에서 제안한 판서 장면 MPEG 동영상에서 문자 인식을 위한 특징을 추출하여 인식하는 방법에 대하여 설명하고, 3장에서는 다음과 모음에 대한 실험 결과 및 고찰에 대하여 기술한다.

2. 시스템 구성

본 장에서는 판서 장면 MPEG Stream Data를 분석하여 문자 인식에 필요한 x, y 좌표 및 방향 코드 등의 특징을 추출하여 HMM 인식기로 문자 인식하는 방법을 설명한다.

그림 1은 본 논문에서 제안한 방법의 전체 흐름도이다. 흐름도는 크게 영상 압축부분, 특징 추출 및 방향 코드 생성부분 그리고 은닉 마르코프 모델 인식기 등의 3부분으로 나눌 수 있다.

본 연구에서 사용한 첫 번째 부분의 영상 압축 방법은 소프트웨어적인 방법으로 Xing사의 XingMPEGEncoder

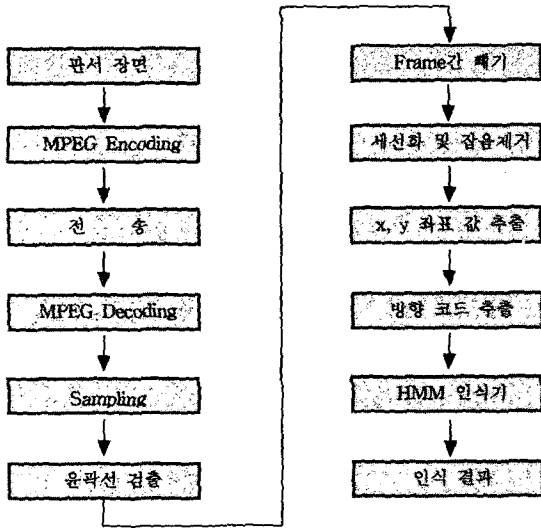


그림 1. 전체 흐름도

를 이용하여, 실제 원격 교육에서 사용하는 환경과 맞추기 위해서 화면의 크기는 320x240으로, Bit rate는 1.2Mbps의 MPEG-1으로 압축하였다[2,3]. 두 번째 부분의 특징 추출 및 방향 코드 생성 과정은 MPEG-1으로 압축된 동영상을 낱장의 Frame으로 분리하여 초당 3 Frame을 Sampling하여 윤곽선 검출, Frame간 빼기 그리고 세션화 및 잡음 제거를 거쳐 x, y 좌표 값과 방향 코드 등의 특징을 추출하는 과정이다[4,5,6,7]. 세 번째 부분은 최종 추출된 8방향 코드와 가상 획 코드를 이용하여 은닉 마르코프 모델 인식기로 인식하는 과정이다. 위의 두 번째와 세 번째 과정에 대해서는 아래에서 좀더 자세히 알아보겠다.

2.1 특징 추출 및 방향 코드 생성

본 절에서는 MPEG-1으로 압축된 동영상에서 특징을 추출하는 방법을 설명한다. 아래는 MPEG 동영상에서의 특징 추출 순서이다.

- ① 전송 받은 동영상을 MPEG Decoder를 이용하여 낱장의 Frame으로 분리한다.
- ② 초당 3 Frame을 Sampling 한다.
- ③ Sampling된 각 Frame에 3x3 ((-1,-1,-1), (-1,-8,-1), (-1,-1,-1)) Laplacian 마스크를 이용한 윤곽선 검출을 한다. 그림 2는 윤곽선 검출의 예이다.
- ④ 순서대로 Frame간 빼기(현재 - 과거)를 한다. 그림 3의 하단부 왼쪽에서 예를 보인다.
- ⑤ 세션화 및 잡음 제거를 한다. 그림 3의 하단부 오른쪽

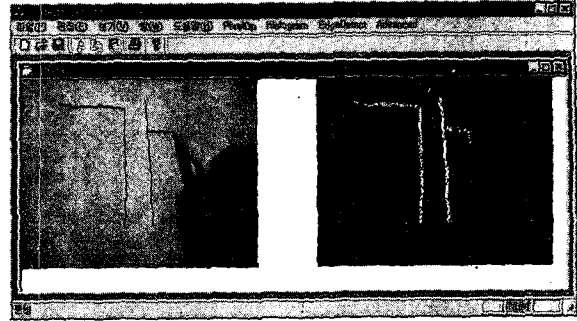


그림 2. 윤곽선 검출

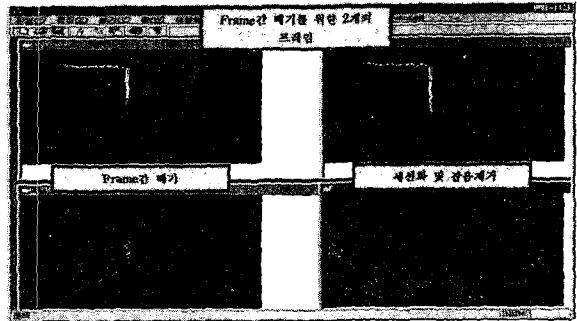


그림 3. Frame간 빼기와 세션화 및 잡음 제거

쪽에서 예를 보인다.

- ⑥ 획에 대한 가로, 세로 중간 Pixel의 x, y좌표 값을 구한다.
- ⑦ 마지막으로, 구해진 x, y좌표 값들로부터 방향코드를 구한다. 그림 4는 방향 코드 및 가상 획 코드를 보인다.

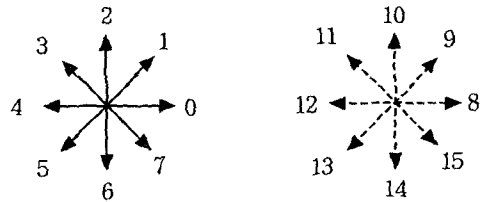


그림4. 8 방향코드 및 가상 획 코드

방향코드는 8방향 코드와 가상 획 코드로 나눌 수 있다. 8방향 코드는 펜의 위치가 X_{old}, Y_{old} 의 좌표에서 새로운 위치로 펜이 이동하는 좌표 값이 X_{new}, Y_{new} 로 되어질 때 x축을 중심으로 각도를 구할 수 있는데, 이 각도를 바탕으로 8 방향코드를 구할 수 있고, 가상 획 코드는 처음 획의 마지막 점과 다음 획의 처음 점을 잇는 가상

의 직선 선분을 가상 획 코드라 정의한다. 이때의 가상 획 코드 값은 8방향 코드 값에 8의 값을 더한 값으로 한다.

2.2 은닉 마르코프 모델 인식기

온라인 상의 필기체 한글문자의 인식은 크게 세 가지의 중요한 요소에 의하여 인식 성능이 좌우된다. 이들 요소는 전처리, 특징추출 그리고 인식 알고리즘의 세 가지이며, 이들 세 가지 요소는 유기적인 관계를 지니고 있다[8]. 본 논문에서 사용하고자 하는 은닉 마르코프 모델은 시간적인 정보를 토대로 특정 사건이 일어날 확률을 예측하는 수학적 모델이므로 온라인 상에서 발생하는 순차적인 문자 데이터를 입력으로 사용하기에 매우 적합하다. 은닉 마르코프 모델은 N개의 전이 상태(state)수와 시간에 따른 전이 상태의 변화를 결정하는 전이 확률 분포(transition probability distribution)A, 각 전이상태에서의 관측 기호들의 확률 분포(observation probability distribution)B, 그리고 A와 B에 대한 초기 상태 확률 분포인 Π 로 구성된다. 본 논문에서 사용한 모델은, 시간에 따라 상태가 일정한 방향으로 진행되는 left-to-right 모델 중에서 simple serial left-to-right 모델로, 그림 5에서 보이고 있다.

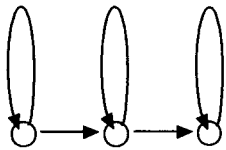


그림 5. simple serial left-to-right 모델

인식 알고리즘인 Viterbi 알고리즘은 관측된 기호들로부터 A와 B의 확률분포로부터 최적의 경로를 추적하게 된다. A와 B의 확률분포는 학습 알고리즘으로부터 얻게 되며 forward-backward 알고리즘, Baum-Welch 알고리즘, Viterbi 알고리즘들이 결합된 형태이다. forward-backward 알고리즘은 확률분포 A, B, Π 에 대하여 관측 심볼들의 집합인 O의 발생확률 계산과정의 중복 계산을 제거하여 시간 복잡도(Time Complexity)를 감축한다. Baum-Welch 알고리즘은 A, B에 대한 재추정을 위한 것으로 학습데이터의 각 문자에 대한 각 경로에서의 누적된 확률 값으로부터 확률분포 A, B가 재추정된다.

3. 실험 결과 및 고찰

본 논문에서 인식률 평가를 위해 사용된 데이터는 한글 자음 5개(ㄴ, ㄷ, ㄱ, ㄴ, ㅎ)와 모음 5개(ㅏ, ㅑ, ㅓ, ㅕ, ㅗ)를 이용하여 실험을 하였다. 표 1에서는 실험 결과를 보인다. 인식 실패의 요인으로서는 글이 써진 부분을 손으로 너무 많이 가렸을 때와 카메라가 심하게 흔들려

잘 못된 x, y 좌표 값을 구했을 때이다.

표 1. 단모음 및 단자음 인식 결과

자음	인식률(%)	모음	인식률(%)
ㄱ	100	ㅏ	80
ㄴ	100	ㅑ	100
ㄷ	80	ㅓ	80
ㄹ	40	ㅕ	100
ㅁ	100	ㅗ	100
ㅂ	100	ㅛ	100
ㅅ	80	ㅜ	100
ㅇ	80	ㅠ	100
ㅈ	60	ㅡ	100
ㅊ	80	ㅣ	100
ㅋ	80		
ㅌ	80		
ㅍ	100		
ㅎ	100		

본 연구에서는 MPEG-1으로 압축된 판서장면 동영상에서 x, y좌표 및 특징을 추출하여 문자 인식하는 방법을 제안하였다. 동영상에서 펜의 움직임에 대한 x, y좌표 값을 구하는 방법은 태블릿을 이용한 다소 제한적인 입력 조건에 의한 문자 인식의 범위를 확장할 수 있으며, 특히 강사의 강의 자료 준비 시간을 없앨 수 있어, 이 같은 경우를 요하는 원격 교육이나 화상 회의에 응용될 수 있으리라 본다.

참고문헌

- [1] 이성환, "문자인식 : 이론과 실제", 홍릉과학출판사, 1993
- [2] Joan L. Mitchell, William B. Pennebaker, Chad E. Fogg & Didier J. LeGall "MPEG Video Compression Standard", Chapman & Hall, 1997
- [3] Prabhat K. Andleigh, Kiran Thakrar, "Multimedia Systems Design", Prentice Hall, Inc., 1996
- [4] 김희승, "영상인식", 도서출판 생능, 1993
- [5] Dana H. Ballard, Christopher M. Brown, "Computer Vision", Prentice Hall, Inc., 1982
- [6] John C. Russ, "The Image Processing Handbook", CRC Press & IEEE Press, 1999
- [7] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Addison-Wesley, Inc, 1992
- [8] C.C Tappert, "Speed, Accuracy, and Flexibility Trade-Offs in On-line Handwriting Recognition", Int. Journal of Pattern Recognition and Artificial Intelligence, Vol.5, No.1&2, pp.79-85, 1991.