음성/영상 연동성능 향상을 위한 입술움직임 영상 추적 테스트 환경 구축

이수종*, 박 준*, 김응규**
*한국전자통신연구원 음성/언어정보연구센터 자동통역연구팀
**한밭대학교 공과대학 정보통신.컴퓨터공학부
e-mail : {sjleetri, junpark}@etri.re.kr, kimeung@hanbat.ac.kr

A Lip Movement Image Tracing Test Environment Build-up for the Speech/Image Interworking Performance Enhancement

Soo-jong Lee*, Jun Park*, Eung-Kyeu Kim**
*Automatic Speech Translation Research Team, Speech/Language Information Research Center, ETRI
**Division of Information Communication & Computer Engineering, Hanbat National University

요 약

본 논문은 로봇과 같이 외부 음향잡음에 노출되어 있는 상황 하에서, 대면하고 있는 사람이 입술을 움직여 발성하는 경우에만 음성인식 기능이 수행되도록 하기 위한 방안의 일환으로, 입술움직임 영상을 보다 정확히 추적하기 위한 테스트 환경 구현에 관한 것이다. 음성구간 검출과정에서 입술움직임 영상 추적결과의 활용여부는 입술움직임을 얼마나 정확하게 추적할 수 있느냐에 달려있다. 이를 위해 영상 프레임율 동적 제어, 칼라/이진영상 변환, 순간 캡쳐, 녹화 및 재생기능을 구현함으로써, 다각적인 방향에서 입술움직임 영상 추적기능을 확인해 볼 수 있도록 하였다. 음성/영상 기능을 연동시킨 결과 약 99.3%의 연동성공율을 보였다.

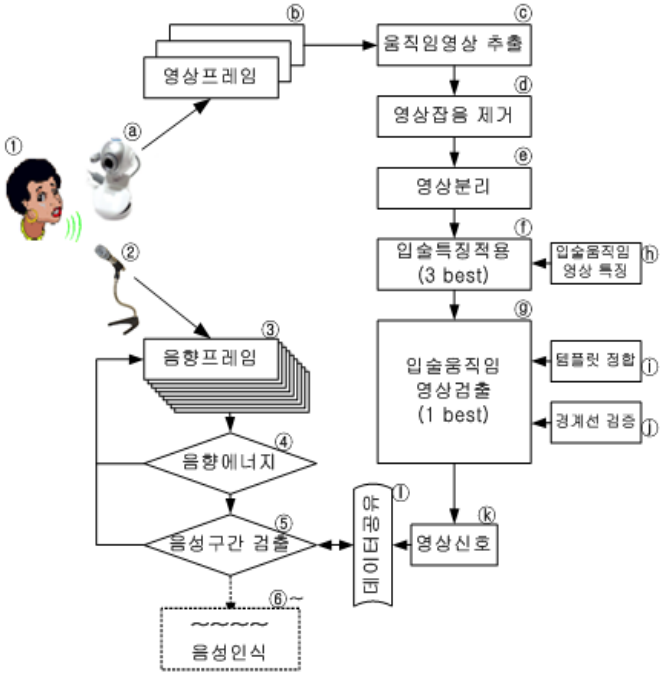
1. 서론

음성인식 기능은 기본적으로 음성에너지를 분석의 대상으로 한다. 그러나, 실제 음성인식 서비스 환경에는 다양한 음향잡음이 존재하게 되어, 이를 제거하기 위한 많은 노력이 기울여지고 있다. 특히, 로봇과 같이 외부 음향잡음에 노출되어 있는 상황에서, 동적잡음이 예고 없이 유입되어 음성으로 오인식 되는 경우에는 심각한 문제를 야기할 수 있다. 한편, 영상의 경우에는 음향잡음과는 무관하게 획득하고 처리될 수 있어서 이를 음성인식에 활용하려는 노력이 계속되고 있다[1][2]. 사람은 말할 때 입술을 움직이게 되므로, 음성인식 과정에 입술움직임 영상 추적 결과를 활용하면 음향잡음을 효과적으로 방지할 수 있게 된다.

본 논문은 음성인식의 전처리 단계인 음성구간 검출 과정에서, 입술움직임 영상추적 결과를 추가 확인할 수 있도록 연동함에 있어서, 입술움직임 영상을 보다 정확히 추적하기 위해 구축한 테스트 환경에 관한 것이다. 입술움직임 추적결과는 영상신호 데이터로 변환되어 최종적으로 공유버퍼(l)에 저장되고 음성구간 검출과정에서 확인하게 된다. 테스트 환경은 Visual C++로 구현되었으며, 영상추적 과정에서 다양한 방법으로 필요한 영상과 데이터를 테스트하고 활용할 수 있도록 구성하였다. 먼저, 음성/영상 연동시스템을 살펴본 다음, 입술움직임 영상 추적에 관한 테스트 환경 구현결과를 서술한다.

2. 음성/영상 연동 시스템 개요

(그림 1)은 기존의 음성인식 절차(②…⑥~)에 영상처리 절차를 부가하여 구축한 구조도이다.

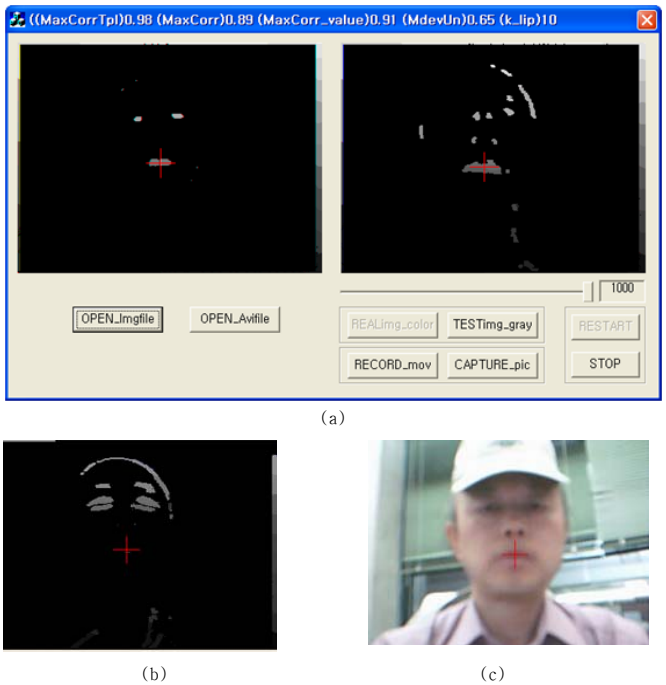


(그림 1) 음성/영상 연동시스템 구조도

음향에너지가 마이크를 통해 입력되어 음성인식 과정에서 분석되는 기존의 방식 외에, PC 용 화상카메라를 통하여 발성화자의 영상을 입력 받아 영상데이터를 동시에 분석하는 구조이다(ⓐ...①)[3][4][5]. 화상카메라를 통하여 움직임이 포착되면 입술움직임 영상이 있는 지의 여부를 분석하게 되고, 분석된 결과는, 마이크에 입력된 음향에너지가 음향잡음인지의 여부를 판정할 수 있게 하는 중요한 신호로 활용된다[6]. 음성인식기와 영상처리기는 간접적으로 연동되는 구조여서, 서로 독립적으로 구동될 수 있다.

3. 입술움직임 영상 추적 테스트 환경

(그림 2)는 음성/영상 연동환경 하에서, 얼굴요소의 움직임 영상 중에서 입술움직임 영상을 추적하는 것을 시각적으로 확인하기 위하여 구축한 테스트 환경(a)과 이를 통하여 추출한 결과의 예(b, c)를 보여준다.



(a)

(b)

(c)

(그림 2) 입술움직임 영상 추적 테스트 화면

(a) 화면의 타이틀 바를 통하여, 입술움직임 영상 추적결과 중 최대적합률(0.98), 현재 추적된 영상의 적합률(0.89) 및 입술움직임 여부를 구분하는 임계적합률(0.65)를 확인할 수 있다. 또한, 영상 프레임율의 동적제어(1~30 프레임/sec)를 통하여 속도를 제어하면서 살펴볼 수 있고, 정지(STOP) 및 재실행(RESTART)할 수 있다. 추적결과를 이진영상(TESTimg_gray, a 의 우측화면)과 칼라영상(REALimg_color, c 화면)으로 살펴볼 수 있다. 필요한 화면을 순간캡쳐(CAPTURE_pic)하여 파일로 저장한 후 열어(OPEN_imgfile, a 의 좌측화면과 b)보면서 확인해 보거나, 동영상으로 저장(RECORD_mov)하여 재생(OPEN_Avifile)해 볼 수 있도록 아이콘으로 구성하였다.

시각적인 확인과 아울러, log 데이터 수집을 통하여 입술움직임 영상 특징데이터를 on-line 으로 추출할 수 있도록 하였다. 다수의 움직임 영상 중에서 입술움직임 영상을 추적하기 위하여는 off-line 에서 미리 입술움직임 특징 파라미터를 설정하게 되는데, 여기에서 추출된 결과를 그대로 활용할 수 있게 되었다. <표 1>은 on-line 으로 수집한 입술움직임 특징 데이터의 예를 보여준다.

<표 1> 입술움직임 특징 on-line 추출결과

특징구분	평균	표준편차
가로폭 (픽셀수)	7	1.29
세로폭 (픽셀수)	27	0.97
가로/세로율	3.86	0.67
면적율	0.42	0.6
명암편차 (픽셀값)	18	0.44
수평 중심율	1.06	1.02
수직 중심율	0.75	0.32

위에서 면적율이란 입술움직임 영상을 포함하는 외접 사각형면적에 대한 실제 면적의 비율이다. 명암편차는 움직임 영상 생성을 위한 영상 프레임간의 명암값 차이를 나타낸다. 중심율은 다수 움직임의 전체 중심으로부터 떨어진 상대적인 거리의 비율이다.

4. 결론

본 논문에서는 음성과 영상의 연동을 통하여 외부의 동적잡음을 효과적으로 방지하기 위한 구조와 함께 입술움직임 영상을 보다 정확히 추적하기 위하여 구축한 테스트 환경에 대하여 살펴보았다. 다양한 방법으로 입술움직임 영상의 추적 상황을 파악할 수 있도록 하였으며, 연동결과 99.3%의 연동성공율을 보였다. 본 연구결과는 연속음성인식 과정에서의 동적잡음 방지에 적극 활용될 수 있을 것으로 기대한다.

참고문헌

[1] G. Potaminanos, H.P. Graf, and E. Cosatto, "An Image Transform Approach for HMM Based Automatic Lipreading, Image Processing, 1988. ICIP 98, Proceeding, pp.173-177, Oct. 1998.

[2] M.T. Chan, Y. Zhang, and T.S. Huang, "Real-Time Lip Tracking and Bimodal Continuous Speech Recognition", IEEE Second Workshop on Multimedia Signal Proceeding, pp.65-70, 7-9 Dec. 1998.

[3] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing, Second Edition", 2002. pp.567-642.

[4] F. Leymarie and M.D. Levine, "Simulating the Grassfire Transform Using and Active Contour Model", Trans. IEEE Pattern Analysis and Machine Intelligence, 14(1):56-75, 1992.

[5] Z.Q.Wu, J.A.Ware, W.R.Stewart, and J.Jiang, "The Removal of Blocking Effects Caused by Partially Overlapped Sub-Block Contrast Enhancement",Journal of Electronic Imaging -- July - September 2005 -- Volume 14, Issue 3, 033006(8 pages).

[6] 이수종, 박준, 이영직, 김응규, “입술움직임 영상신호를 고려한 음성존재 검출”, 한국음향학회지 제 26 권 제 1 호, 2007.1, pp.25-31.