

입술 움직임 영상 신호를 이용한 음성 구간 검출

Speech Activity Detection using Lip Movement Image Signals

김응규*
Eung-kyeu Kim*

요약

본 논문에서는 음성인식을 위한 음성구간 검출과정에서 유입될 수 있는 동적인 음향에너지 이외에 화자의 입술움직임 영상신호까지 확인함으로써 외부 음향잡음이 음성인식 대상으로 오인식되는 것을 방지하기 위한 한 가지 방법이 제시된다. 우선, 연속적인 영상이 PC용 영상카메라를 통하여 획득되고 그 입술움직임 여부가 식별된다. 다음으로, 입술움직임 영상신호 데이터는 공유메모리에 저장되어 음성인식 프로세서와 공유한다. 한편, 음성인식의 전처리 단계인 음성구간 검출과정에서 공유메모리에 저장되어진 데이터를 확인함으로써 화자의 발성에 의한 음향에너지인지의 여부가 입증된다. 최종적으로, 음성인식기와 영상처리를 연동시켜 실험한 결과, 영상카메라에 대면해서 발성하면 음성인식 결과의 출력에 이르기까지 연동처리가 정상적으로 진행됨을 확인하였고, 영상카메라에 대면치 않고 발성하면 연동처리시스템이 그 음성인식 결과를 출력치 못함을 확인하였다. 또한, 오프라인하의 입술움직임 초기 특징값 및 템플릿 초기영상을 온라인 하에서 추출된 입술움직임 초기특징값 및 템플릿 영상으로 대체함으로써 입술움직임 영상추적의 변별력을 향상시켰다. 입술움직임 영상 추적과정을 시각적으로 확인하고 실시간으로 관련된 페러미터를 해석하기 위해 영상처리 테스트베드를 구축하였다. 음성과 영상처리 시스템의 연동결과 다양한 조명환경 하에서도 약 99.3%의 연동율을 나타냈다.

Abstract

In this paper, A method to prevent the external acoustic noise from being misrecognized as the speech recognition object is presented in the speech activity detection process for the speech recognition. Also this paper confirmed besides the acoustic energy to the lip movement image signals. First of all, the successive images are obtained through the image camera for personal computer and the lip movement whether or not is discriminated. The next, the lip movement image signal data is stored in the shared memory and shares with the speech recognition process. In the mean time, the acoustic energy whether or not by the utterance of a speaker is verified by confirming data stored in the shared memory in the speech activity detection process which is the preprocess phase of the speech recognition. Finally, as a experimental result of linking the speech recognition processor and the image processor, it is confirmed to be normal progression to the output of the speech recognition result if face to the image camera and speak. On the other hand, it is confirmed not to the output the result of the speech recognition if does not face to the image camera and speak. Also, the initial feature values under off-line are replaced by them. Similarly, the initial template image captured while off-line is replaced with a template image captured under on-line, so the discrimination of the lip movement image tracking is raised. An image processing test bed was implemented to confirm the lip movement image tracking process visually and to analyze the related parameters on a real-time basis. As a result of linking the speech and image processing system, the interworking rate shows 99.3% in the various illumination environments.

Keywords : Lip movement image signals, Acoustic noises, Speech & image processing system, Speech recognitions

I. 서론

음성인식은 기본적으로 음향에너지를 분석의 대상으로 한다. 그런데, 음성인식의 실제 서비스 환경에서 마이크를

통해 음성인식 과정에 유입되는 음향에너지 내에 배경잡음 및 다양한 동적 음향잡음이 존재한다[1-2]. 특히, 음성인식

로봇과 같이 외부 음향잡음에 노출되어 있는 상황에서 음성인식을 수행하는 경우 음성인식 대상이 아닌 동적잡음이 예고 없이 유입될 수 있으며 이들 잡음이 음성으로 오인식되는 경우 심각한 문제를 일으킬 수 있다. 한편, 사람은 말을 하거나 대화할 때 입술을 움직이게 된다. 입술을 움직이

* 한밭대학교

투고일 : 2010. 7. 7 수정완료일 : 2010. 10. 20

게재확정일 : 2010. 10. 29

지 않고 말을 할 수 있는 방법은 거의 없다. 이에 반해 영상은 음향잡음과 무관하게 획득되고 처리될 수 있어 음성인식을 위한 음향에너지 분석과정에서 입술움직임 영상신호의 도입은 음향잡음 처리를 위한 효과적인 대책이 될 수 있어, 이를 음성인식에 활용하려는 노력이 계속되어 왔다 [3-4]. 이에, 이러한 영상의 장점을 음성인식에 활용하기 위한 독순술(lipreading)에 관한 연구가 꾸준히 진행되어 왔다 [5]-[7]. 하지만, 주로 극한 소음환경하에서 입술모양만으로 음성인식을 시도하는 경우와 입 모양의 구분이 명확한 모습과 일부 단어를 인식하는 수준에 머물고 있는 경우가 많았다 [8]-[10]. 이는 영상과 음성신호자체의 속성 차이 및 영상신호 처리량의 증가 때문인 것으로 풀이된다. 따라서, 음성정보와 영상정보를 보다 효율적으로 결합함으로써 기존의 음성인식 속도와 인식률을 저하시키지 않으면서 음성인식 대상이 아닌 외부 음향잡음을 효과적으로 차단하고, 음성인식 절차가 불필요하게 진행되는 것을 방지할 필요성이 제기되었다.

본 논문은 외부의 음향잡음이 음성인식 대상으로 오인식되는 것을 방지하기 위한 방안의 하나로 행해졌으며, 영상처리와 음성인식이 독립적으로 수행되도록 함으로써 기존 음성인식 과정에서 추가될 수 있는 부하량이 최소화 되도록 하였다.

우선, 음성/영상 연동 시스템을 구축하여 음성인식의 전처리 단계인 음성구간 검출 과정에서 음향에너지에 대한 기존의 분석 이외에 화자의 입술움직임 영상신호까지 추가로 확인하는 등 다양한 영상환경에서도 입술움직임 영상신호를 보다 정확히 확인하는 구조로서 Visual C++로 구현하였다 [11]-[16]. 아울러, 음성구간 검출과정에서 입술움직임 영상신호의 유무에 따라 음성인식 절차로의 진행여부가 결정되도록 하였다. 이를 위해 온라인 하에서 입술움직임 표준 영상을 추출하였고, 이에 관한 특징벡터를 추출하여 오프라인에서 산출된 특징벡터 초기 모델과 템플릿 초기 모델을 대체하였다. 그리고, 입술움직임 영역 추적에 관한 주요 파라미터를 시각적으로 확인하고 검증하기 위하여 영상처리 테스트베드를 구축하였다.

이하, 2장에서는 음성 인식용 시스템의 개요에 대해 개략적으로 살펴본다. 3장에서는 입술움직임 영상획득에서 영상신호 추출까지의 여러 단계와 영상처리 테스트베드 구축 및 시스템 연동에 대하여 상세히 기술하고, 4장에서는 실험환경을 포함한 음성/영상 연동 실험결과에 대해 기술한다. 5장에서는 결론 및 향후과제에 대해 기술한다

II. 음성 인식용 시스템의 개요

그림 1은 음성인식과정에서 영상정보를 활용하는 음성/영상결합 음성인식시스템의 개요를 나타낸 것이다. 입술움직임 영상신호의 추출기능과 음성인식의 모든 과정은 하나의 PC 내에서 윈도우가 분리된 인터페이스를 통해 이루어진다. 음성인식을 위해 화자가 발성함에 따라 마이크를 통해서 음향에너지가 수신되어 분석되며, 동시에 영상카메라를 통해서 얼굴영상 프레임이 연속적으로 획득되고 입술움

직임 영상을 검출한다. 입술움직임 영상분석 데이터를 음성인식 과정에서 공유하도록 하기 위하여 공유메모리가 준비된다. 음성인식의 음성구간 검출과정(speech activity detection)에서는 음향에너지의 분석과 함께 공유메모리로부터 입술움직임 영상신호를 확인하여 음성구간 검출을 수행하고 음성인식 절차로의 진행여부를 결정하게 된다.

입술움직임 영상신호의 추출을 위해서는 실제 더 많은 처리절차를 거치게 된다. 즉, 영상 프레임간 비교, 잡음영상 제거, 영상분리, 입술움직임 영상 특징추출 및 검출, 영상신호로의 변환, 공유메모리내 저장 등이다. 이들에 대해서는 다음 장에서 상세히 기술한다.

한편, 음성인식의 음성구간 검출기능에는 입술움직임 영상신호를 확인하는 절차를 추가하게 된다. 이러한 영상신호 확인기능 확대에 따른 추가될 수 있는 부하를 최소화하는 것이 필요하다. 이에 대한 언급은 입술움직임 영상데이터 확인절차를 포함하여 구동되는 음성/영상 시스템의 연동절차를 서술하는 과정에서 이루어진다. 그리고, 음성인식 결과의 출력여부를 통한 시스템의 정상적인 동작여부에 대해 다음과 같이 확인한다. 즉, 카메라로 하여금 입술움직임을 감지하는 환경에서 발성하는 경우 음성인식을 수행한다. 한편, 카메라로 하여금 입술움직임 영상을 획득하지 못하도록 한 상태에서 발성한다고 해도 인식결과가 출력되지 않아야 한다. 그 이유는 사람이 입술을 움직여 발성한 경우외에 외부 음향잡음이 음성으로 오인식되어 발생되지 않도록 하기 위함이다.

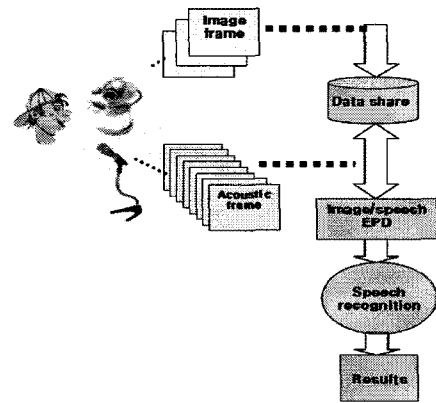


그림 1. 음성/영상결합 음성인식시스템의 개요
Fig. 1. Overview of speech recognition system for combining speech and image

그림 2는 그림 1을 상세히 나타낸 음성/영상 결합시스템의 구조이다. 녹색부분의 음성인식기와 노란색 부분의 영상처리기가 공유메모리(k)를 통하여 연동되며 영상처리가 진행되는 시점부터 입술움직임 영역의 추적결과가 공유메모리에 저장되기 시작하고 음성인식의 음성구간 검출과정에 반영된다.

이 연동시스템에서의 음성인식 절차를 요약하면 다음과 같다. 우선, 음성인식을 위한 사람의 발성 뿐만 아니라 다양한 형태의 외부 음향잡음을 포함한 음향에너지가 마이크

(1)을 통하여 입력된다. 이때 음향흐름의 데이터는 적절한 크기의 프레임별(2)로 나누어져 필터링 과정을 거친 후 일정한 크기 및 고주파를 특징으로 하는 정적인 잡음이 제거되어진다(3). 다음으로, 음향에너지의 크기와 지속성 및 새로 추가된 입술 움직임 영상신호 데이터(i)를 확인한 후 공유메모리를 거쳐 음성구간으로 표시한다(4). 그 다음으로, 음성구간을 대상으로 프레임별로 특징값을 (5)하고 음성인식(6)을 수행한다. 이때 미리 구축된 음성/언어 데이터베이스(10)를 기반으로 하는 음향모델(7), 발음사전(8), 언어모델(9)들을 통해 음성인식을 위한 탐색공간이 형성된다. 이 음성인식(6) 과정에서는 탐색공간 내의 음성 특징값(5)을 비교하여 인식결과(11)를 도출한다.

모리(k)에 저장되고 음성구간 검출과정(4)에서 활용된다.

III. 입술움직임 영상신호 추출 및 시스템 연동

입술움직임 영상신호를 음성인식 구간검출에 활용하기 위해서는 카메라를 통하여 연속적으로 입력되는 영상프레임을 대상으로 영상배경, 얼굴전체의 움직임 및 얼굴요소의 움직임 영상으로부터 입술움직임 영상만을 분리해내야 한다. 이를 위해 연속되는 인접 영상프레임을 픽셀단위로 비교하는 것에서 시작하여 여러 단계를 거쳐게 된다. 그림 3은 영상 프레임으로부터 영상신호 추출에 이르기까지의 단계를 나타내며, 아울러 이에 따른 주요 추출단계의 각 항목을 기술한다. 그리고 입술움직임 추적과정을 시각적으로 확인하기 위한 영상처리 테스트베드 구축과 음성/영상 시스템의 연동에 대해서도 기술한다.

3.1 연속영상 프레임 획득과 움직임 영상 검출

우선, 연속영상 프레임 획득과 관련하여 기술하면, PC용 영상카메라는 초당 30 프레임의 영상을 획득한다. 실제 프로그램 진행과정에서 부하에 의해 그 속도가 저하하게 된다. 그럼에도 불구하고 최소 15 프레임의 속도를 유지할 수 있어 연구를 수행함에 있어 큰 지장은 없다. 화면의 크기는 가로 픽셀수와 세로 픽셀수가 320×240으로, 칼라 영상 전체 화면을 처리하기 위해 1회에 20,400번의 처리가 요구된다. 따라서 실행속도의 향상을 위해 컬러와 흑백 영상으로의 적절한 변환과 부분적인 처리 방법 등을 강구하였다.

다음으로, 움직임 영상 검출에 관해 기술하면, 일단, 입술 움직임이 있으면 두 영상 프레임 간에 변화가 있는 부분이 존재하게 된다. 또한, 영상 프레임의 동일 위치에 대한 픽셀값의 차이를 분석하면 움직임 영역을 찾아낼 수 있다. RGB 컬러모델로부터 HSI 컬러모델의 명암값, $I = (R + G + B) / 3$ 로 변환한 다음, 각 픽셀의 명암값 변화정도에 따른 움직임 영상 검출의 관계는 다음 식(1)로 나타낼 수 있다.

$$d_{ij}(x, y) = \begin{cases} 255 & \text{if } |f(x, y, t_i) - f(x, y, t_j)| > T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

여기서, 시간 t_i 와 시간 t_j 에서 획득한 두 영상 프레임 $f(x, y, t_i)$ 와 $f(x, y, t_j)$ 간의 비교 결과인 $d_{ij}(x, y)$ 는 공간좌표 (x, y) 에서 두 영역간 픽셀값의 차이가 문턱값 T 보다 클 경우에만 8bits로 표시된 명암값 255를 갖도록 하였다 [17]. 조명의 미세한 변동과 같이 실제 움직임과는 무관하게 나타날 수 있는 영상잡음의 제거를 위해 문턱값의 설정이 요구된다.

구현에 있어서 문턱값을 10으로 하되, 명암값 차이의 정도를 입술움직임 영상의 특징값으로 활용하기 위해 255 대신 $|f(x, y, t_i) - f(x, y, t_j)|$ 의 값을 그대로 산출하여 중요한 특징요소의 하나로 활용하였다. 문턱값 이상의 움직임 영상의 경우에도 다양한 요인에 의하여 미세한 크기의 잡음영상이 다수

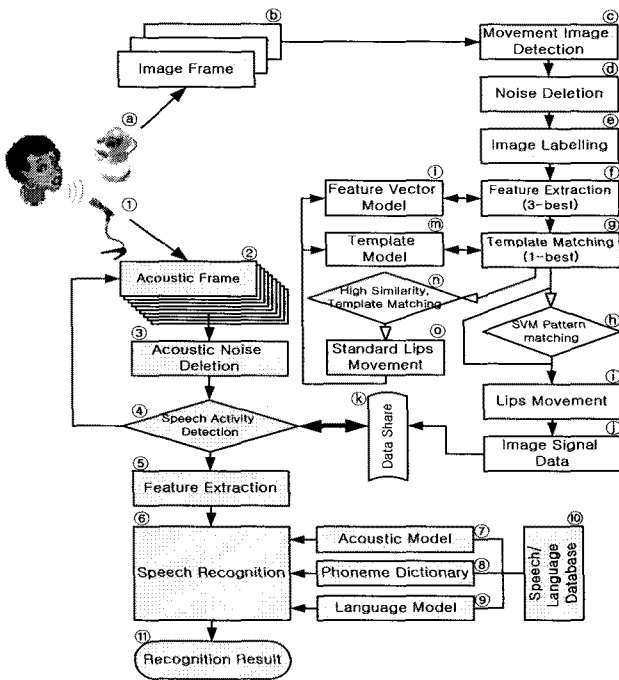


그림 2. 음성/영상 결합시스템의 상세구조

Fig. 2. Detail structure of the combined speech and image system

한편, 음성인식을 위해 발생하는 화자의 움직임 영상은 PC용 영상카메라(a)를 통해 영상프레임별(b)로 획득된다. 인접한 영상 프레임간의 비교에 의한 움직임 영역(c) 추출, 잡음영역 제거(d), 영역별 라벨링(e), 라벨링 영역별 특징추출(f)과 특징벡터 모델(1)에 의한 입술움직임 후보영역 선정 및 템플릿 모델(m)에 의한 최종후보를 선정한다(g). 또한 템플릿 모델에 의한 템플릿 적합도(혹은 정합률, fitness rate)가 산출되고 템플릿 적합도 분포에 의해 입술움직임 영역 또는 여타의 얼굴요소 영역을 구분하는 임계값이 형성된다. 입술 움직임 최종후보는 템플릿 적합도가 임계값 이상인지의 여부에 따라 결정된다. 그러나 템플릿 적합도의 측정이 미흡한 초기단계에서는 SVM (Support Vector Machine)에 의한 영역분류(h)를 거친 후 실제 입술움직임 인지(i)의 여부가 판별된다. 템플릿 적합도는 영상신호로 변환되어 공유메

포함되는데, 가로 5×세로 7 크기 이하의 영상은 입술움직임 영상의 크기를 감안하여 제거되었다[17].

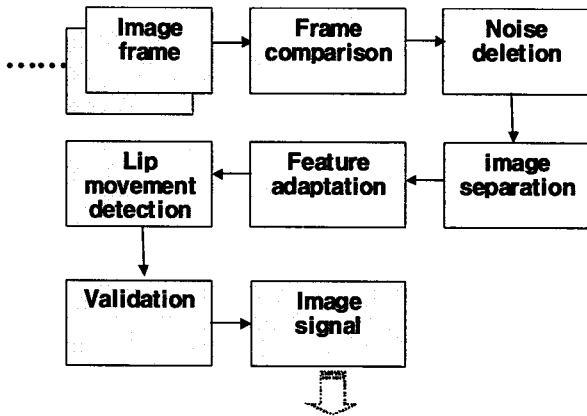


그림 3. 입술움직임 영상신호 추출단계

Fig. 3. Detection procedure of lip movement image signal

3.2 움직임 영상분리 및 입술움직임 영상특징 적용

움직임 영상은 여러 부분에서 나타날 수 있다. 몸의 움직임, 얼굴 움직임, 눈의 깜빡임, 턱의 움직임 등이다. 이들 각각은 식별을 위해 분리하여 개별화해야 한다. 따라서 움직임 영상의 각 각을 라벨링(labeling)하여 영역별로 분리하였다. 영상분리는 그래스파이어 변환(grassfire transform)[18]을 적용하여 분리된 각각의 움직임 영역을 대상으로 입술움직임 영상 특징과 비교하였다. 그리고, 입술움직임 영상 특징과의 유사도(similarity)가 높은 순서에 따라 일부 움직임 영상만을 대상으로 다음 단계의 세밀한 적합도(fitness rate) 산출 단계를 거쳐 입술움직임 영상 여부를 판별하였다.

또한, 입술움직임 영상특징 요소를 선정함에 있어서 눈의 움직임 영상과 대비될 수 있는 파라미터를 발굴하였다. 왜냐하면, 눈의 깜빡임은 인위적으로 조절할 수 없는 요소일 뿐만 아니라 입술움직임 영상의 특징과 유사성이 많기 때문이다. 입술움직임 특징요소별 데이터인 입술움직임 영역 특징벡터(1)은 다수의 움직임 중에서 입술움직임 영역과 매우 유사한 3개의 후보(f)를 선정하는 기준이 된다. 입술움직임 영역의 특징요소로서 그 폭과 높이, 이들의 비율, 넓이, 평균 픽셀 값 차이, 공간 좌표상의 상대적 위치 등 7가지를 설정하였다. 입술움직임 영역 특징벡터(1)의 초기화에 대비하여 오프라인하에서 미리 이들 데이터를 수집하기 위해 카메라로부터 50cm 거리에서 나타나는 움직임 영상을 대상으로 수집하여 초기값으로 활용하였다. 영상은 조명과 같은 영상 획득환경에 따라 특징값에 대한 편차가 많지만 상대적인 유사도 산출에 활용 가능하다. 입술움직임 영상의 주요 파라미터 및 특징값을 표 1에 요약하였다.

표1에 표시된 숫자는 입술움직임 영상을 구성하는 픽셀의 수 또는 이들의 비율과 각 요소값의 표준편차를 나타낸다. 면적률(dimension rate)은 입술움직임 영상의 면적을 외접사각형의 넓이로 나눈 것이다.

표 1. 입술움직임 영상 특징 요소별 데이터

Table 1. Data of each feature elements of lip movement images

Lip Feature Vector	(1) length	(2) width	(3) width/leng.	(4) dimension rate	(5) pixel value	(6) length location	(7) width location
average	5	20	3.95	0.37	16	1.03	0.68
stand. dev.	2.18	6.88	0.87	0.07	3.92	0.18	0.1

위에 표시한 숫자는 입술영상을 구성하는 픽셀의 개수 또는 이들의 비율 및 각 요소값의 표준편차를 나타낸다. 면적률(dimension rate)은 입술움직임 영상의 면적을 외접사각형의 넓이로 나눈 것이다. 평균 픽셀값(pixel value)은 프레임간의 비교를 통하여 움직임 영상을 검출하는 과정에서 산출된 명암값 차이의 평균이다. 가로 및 세로위치(length and width location)는 다수의 움직임 영상들의 중심좌표로부터 떨어진 정도이다. 구현을 위해 입술움직임 영상의 특징 요소외에도 좌우상하 특성으로서 해당 방향으로 분포한 움직임 영상의 분포정도를 부가하여 적용하였다. 예를 들어, 눈의 움직임 영상의 경우 좌우에 최소한 다른 눈의 움직임 영상이 있고, 입술의 경우 아래 부분에 턱의 움직임 영상이 있다. 입술움직임과 관련하여 여타의 움직임과 차별화할 수 있는 특징 파라미터의 지속적인 발굴이 필요하다.

입술움직임 영상의 특징값은 다수의 움직임 영상들의 특징값과의 비교를 위해 기준값으로 사용된다. 그러나 개별적인 특징요소별로 입술움직임과 영상획득 환경에 따라 데이터의 변화가 많기 때문에 특징벡터들을 종합하여 유사성의 척도로 활용하였다. 많은 움직임 영상(M_i)을 대상으로 입술움직임 영상(L_i) 특징벡터와 대비시킨 유사성 척도($M_i sim$)를 다음 식 (2)와 같이 산출하였다.

$$(M_i) sim = \sum_{j=1}^k \left(-0.1 * \frac{|M_{i(j)} - L_i avg_{(j)}|}{L_i std_{(j)}} + 1 \right) * w_{(j)} \quad (2)$$

식 (2)는 움직임 영상(M_i)의 특징요소(j)의 특징값과 입술움직임 영상 특징요소(j)의 기준값($L_i avg_{(j)}$)간의 차이의 절대값을 입술움직임 영상 특징요소(j)의 표준편차($L_i std_{(j)}$)로 나눈 것이다. 즉, 움직임 영상의 유사도(similarity)가 1.0 이하의 값이 되도록 하였고 입술움직임 영상 특징과의 편차가 클수록 낮은 값이 되도록 하기 위해 기울기 -0.1과 절편 1을 설정하였다. 특징요소별 가중값($w_{(j)}$)은 별도로 설정하지 않고 균등하게 $1/k$ 로 적용하였다. 움직임 영상 프레임마다 다수의 움직임 영상을 대상으로 입술움직임 영상 특징벡터와의 유사도를 산출하였고, 이 결과를 토대로 각각의 움직임 영상 프레임별로 유사도가 높은 순으로 6개씩의 움직임 영상을 선정하였다. 프레임별로 유사도가 높은 순으로 선정한 6개씩의 움직임 영상의 예를 표 2에 E로 적표 2에서 F_j 는 움직임 영상 프레임을 나타내고, M_k 는 움직임 영상 프레임 내에 있는 움직임 영역들을 나타낸다. 프레임 내에는 6개 미만의 움직임 영역이

있는 경우(F_1, F_2, F_9)도 있는데, 6개 이상인 경우가 대부분이다.

표 2. 움직임 영상 프레임과 그 움직임 영역의 유사도 예
Table 2. Examples of similarities between movement image frames and its movement regions

	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	...
F ₁	0.38	0.06	0	0	0	0	
F ₂	0.54	0.40	0.38	0.28	0.18	0	...
F ₃	0.68	0.58	0.58	0.45	0.29	0.28	...
F ₄	0.93	0.75	0.66	0.56	0.31	0.29	...
F ₅	0.94	0.84	0.8	0.55	0.51	0.40	...
F ₆	0.94	0.74	0.64	0.51	0.40	0.29	...
F ₇	0.84	0.76	0.72	0.56	0.29	0.23	...
F ₈	0.93	0.82	0.77	0.77	0.70	0.33	...
F ₉	0.82	0.72	0.50	0.30	0.28	0	...
F ₁₀	0.47	0.39	0.32	0.29	0.29	0.27	...
...

여기에서는 최대 6개까지 선정된 움직임 영역을 대상으로 입술 움직임 영상특징과의 유사도 예를 나타냈다. 입술 움직임 영상특징과의 유사도가 가장 높은 영역(M₁)으로부터 가장 낮은 영역(M₆)에 이르기까지 차례로 나타냈다. 프레임에 따라서 유사도가 모두 낮아 입술움직임 영상이 포함되지 않았다고 판단되는 경우도 있고, 0.90 이상의 유사도를 갖는 영역(F₄, F₅, F₆, F₈)도 있음을 나타내고 있다. 또한 입술움직임 영역에 대한 포함여부의 판단이 어려운 경우도 있었다(F₂, F₃). 따라서 입술움직임 영상 여부를 판단하기 위하여 유사도외에 추가적인 분석이 필요하다.

3.3 템플릿 적합도 산출 및 입술움직임 영상 검출

움직임 영상 중에서 입술움직임 영상특징과의 유사도가 상대적으로 높은 경우에 실제 입술움직임 영상인지의 여부를 판별하기 위한 절차가 필요하다. 이는 입술움직임 외에도 움직임 요소가 많기 때문이다. 또한 입술움직임 정도와 경우에 따라 입술움직임 영상의 크기와 모양이 다양하게 변화하게 된다. 이런 점들을 감안하면, 입술움직임 영상특징 외에 입술움직임 영상 자체와의 직접적인 비교만이 가장 확실한 방안이 될 수 있다. 하지만 영상간의 직접적인 비교는 많은 계산량이 요구된다는 점과 입술영상 자체는 움직임에 따라 변화가 많다는 점을 감안해야 한다. 이에 대해 여기에서는 입술움직임 영상특징과의 유사도가 높은 순으로 3개의 움직임 영상만을 대상으로 입술 주위의 일부 영역으로 한정하여 비교대상 템플릿을 정하여 움직임 영상과의 적합도(fitness rate)를 분석하는 방법을 택하였다. 이때, 이 3개의 입술움직임 후보 영상 및 그림 2(㉠)의 특징벡터 초기모델과 최종후보를 선정하기 위한 그림 2(㉡)의 템플릿 초기 모델은 오프라인 하의 특정한 영상 환경에서 산출된 것이다. 다양한 영상획득 환경으로의 적용을 위해 온라인 하에서 입술움직임 표준영상을 추출한 후 이를 매개로 하여 오프라인 하에서 산출된 특징벡터 초기모델과 템플릿 초기모델을 온라인 하에서 추출한 특징모델과 템플릿 모델

로 각각 대체하였다. 3개의 입술움직임 영상들에 대한 적합도를 산출하였고, 가장 높은 적합도를 보이는 움직임 영역을 대상으로 입술움직임 영상여부를 판별하도록 하였다. 이를 위해 적합도에 대한 임계값 설정이 필요하다.

적합도 산출을 위한 템플릿은 그 크기와 모양의 변화가 거의 없고 명암의 대비 혹은 구분이 뚜렷한 코의 일부 영상으로 하였다. 왜냐면 템플릿 적합도는 픽셀단위로 픽셀값 사이의 편차를 구함으로써 영상간의 상관성을 보다 정확히 측정하는데 적합한 방법이기 때문이다. 코를 포함하는 얼굴 영상 파일을 저장해 놓고 적합도 산출과정에서 이를 읽어 들인 후, 코의 일부 템플릿 영역과 입력되는 움직임 영상의 중심으로부터 위쪽으로의 일정부분이 비교되도록 하였다. 즉 움직임 영역이 입술일 경우에는 상단 일정부분에는 코의 영상이 있을 것이고 코의 템플릿과의 비교를 통해 적합도가 높게 나오게 된다. 적합도 데이터가 축적됨에 따라 입술움직임 영상과의 적합도 분포범위가 여타 움직임 요소의 적합도와 구분되도록 하였다. 그림 4는 움직임 영상의 템플릿 적합도와 그 빈도수의 분포 예를 나타낸 것이다.

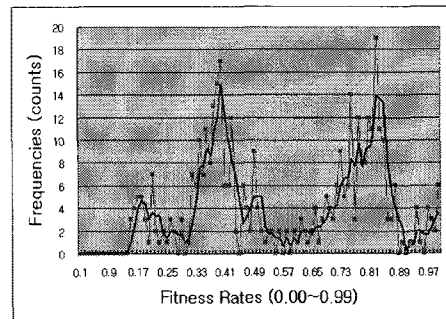


그림 4. 움직임 영상들의 템플릿 적합도 분포
Fig. 4. Distribution of template fitness rates of movement images

템플릿 적합도에 따른 빈도 수의 편차가 많으므로 3 구간 이동평균선인 검은 곡선을 구하여 분석에 활용하였다. 우측의 블록곡선은 입술움직임 영상의 적합도이고 좌측의 블록곡선들은 여타 얼굴요소의 움직임 영상에 대한 것임을 실험을 통해 확인하였다. 첫 번째 블록곡선과 두 번째 블록곡선 사이의 계곡점 즉, 여기에서 0.57이 입술영역을 구분하는 임계값이 된다. 임계값이 자동으로 형성되고 검출되는 과정에서 관련 데이터 축적을 위해 실행 초기에 몇 초 정도의 시간이 소요될 수 있다.

이상의 절차를 그림 5에 요약했다. 즉, 프레임 별 움직임 영상의 특징들과 미리 준비된 입술움직임 영상특징을 비교하여 유사도가 높은 순으로 3개의 움직임 영상을 검출한다. 이 3개의 움직임 후보영상(f)을 대상으로 템플릿 모델영상(m)과의 픽셀단위에 의한 직접 비교를 통해 적합도가 가장 높은 1개의 최종후보 움직임 영상(g)을 검출한다. 적합도별로 빈도를 분석한 후에 입술움직임 영상 판별을 위한 임계값을 산출한다. 임계값 보다 큰 적합도를 보이면 입술움직임 영상으로 분류한다.

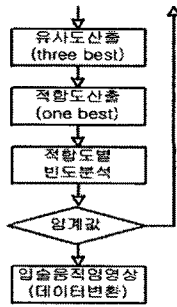


그림 5. 입술움직임 영상의 검출절차

Fig. 5. Detection procedure of lip movement images

최종적으로, 이전의 검출과정에서 높은 적합도를 나타냈던 위치를 중심으로 2단계 입술움직임 영상의 존재여부를 확인한다. 왜냐하면, 입술움직임의 크기와 모양이 다양하게 변하는 점에 의해 실제 입술움직임의 영상인 경우에도 입술움직임 영상특징과의 유사도가 다소 낮아질 수 있어 이러한 영상들도 가능한 찾아내야만 하기 때문이다. 따라서 임계값 이상의 적합도를 보이는 움직임 영상이 검출되지 않는 경우에 2단계로 재검출하는 방법을 택하였다.

3.2절 내의 유사도 산출결과 중에서 활용하지 않았던 움직임 영상을 포함하여 모든 움직임 영역인 $M_1 \sim M_6$ 을 확인하였다. 단, 2단계 입술움직임 영상을 검출하기 위해 얼굴 전체를 과도하게 움직이지 않는다는 전제조건이 필요하다.

그외에 여기에서 사용된 템플릿 적합도를 측정하는 모형과 산출식을 다음 그림 6에 나타낸다. 그림 6(b)에서 $a(x,y)$ 는 입력영상에서 취한 비교부 $g(x,y)$ 에서 평균 $E(g)$ 를 뺀 밝기값을 나타내며, $b(x,y)$ 는 템플릿 영상 $t(x,y)$ 에서 평균 밝기 $E(t)$ 를 뺀 밝기값을 나타낸다. x, y 는 수직 및 수평방향으로의 인덱스로서 픽셀좌표이다. 적합도 산출식(b)는 농담정규화 상관법 (Normalized Gray-level Correlation)으로서, 템플릿(그림 6(a)-⑦)과 이와 비교할 검색범위의 일정영역(그림 6(a)-①,②,⑤)을 밝기에 대해 정규화시켜 비교함으로써 전체적인 조명의 밝기가 높아지거나 낮아지는 경우에도 패턴의 비교를 가능하게 한다. 검색범위는 각 움직임 후보(그림 6(a)-③,④,⑥)의 바로 위에 위치한다. 템플릿 적합도가 산출되고 축적되면서 입술움직임 여부를 구분할 수 있는 임계값이 형성된다.

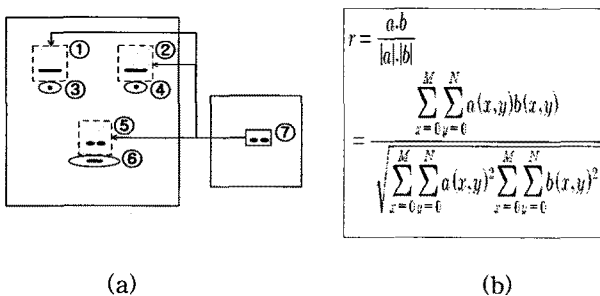


그림 6. 템플릿 적합도 측정 모형(a)과 산출식(b).

Fig. 6. Measuring model of template matching(a), and calculating formula(b)

3.4 명암값 보상 및 입술움직임 영상신호 변환

우선, 명암값 보상에 대하여 기술하면, 움직임 영상과 템플릿 영상의 픽셀 단위 비교시 주변의 조명에 의하여 적합도 산출은 많은 영향을 받게 된다. 이를 보상하기 위하여, 주변의 조명에 무관하게 명암값이 0~255 사이에 전반적으로 분포하도록 명암 분포를 확장시켰다[17]. 템플릿 영상과 움직임 영상에서 적합도 대상 부분에 한정하여 적용하였다. 명암 대비 확장(stretching)은 다음과 같은 매핑함수로 표현된다.

$$V_{new}(x,y) = \frac{V_{old}(x,y) - V_{min}}{V_{max} - V_{min}} * 255 \quad (3)$$

여기에서 $V_{old}(x,y)$ 는 템플릿 영상을 포함하여 실시간으로 입력되는 움직임 영상의 픽셀값을 나타내고, $V_{new}(x,y)$ 는 변환된 값을 나타낸다. V_{min} 은 입력 영상의 픽셀값 중에서 최소값을 나타내고 V_{max} 는 최대값을 나타낸다.

다음으로, 입술움직임 영상신호 변환에 대해 기술한다. 즉, 입술움직임 영상 특징과의 유사도가 높고 임계값 이상의 템플릿 적합도를 가진 입술움직임 영상이 검출되면 이를 음성구간 검출과정에서 활용할 수 있도록 데이터로 변환해야만 한다. 영상 신호값은 입술움직임 영상 특징과의 유사도와 템플릿 적합도를 합산하여 구했다. 영상 신호값은 0.0~1.0 사이의 값을 갖도록 하였고, 음성인식의 음성구간 검출모듈에서는 그 값이 양의 값을 가지면 입술움직임이 있는 것으로 판단하도록 하였다.

3.5 영상처리 테스트베드 구축

입술움직임 영역 추적 과정과 오류 여부를 시각적으로 확인하고 관련 파라미터를 실시간 분석하기 위하여 영상처리 테스트베드를 구축하였다. 그림 7은 발생목록에 따라 발생하면서 입술움직임 영역을 추적하는 테스트베드 구축상태를 나타낸 것으로, 음성인식 결과의 확인에 따른 입술움직임 영역추적이 잘 이루어지고 있음을 시각적으로 알 수 있다. 그리고, 인터페이스는 크게 세 부분으로 나뉘어지는데, 파라미터 확인부분, 영상 제어부분 및 영상 재생부분이다.

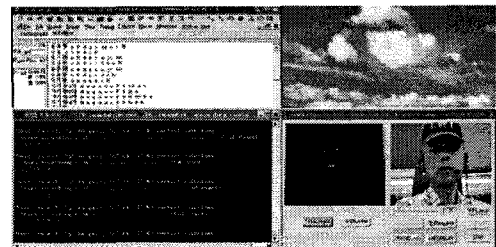


그림 7. 입술움직임 영역 추적 테스트베드.

Fig. 7. Test bed for tracking lips movement region

우선, GUI 구성에 대해 기술하면, 타이틀 바에 주요한 파라미터가 나타나도록 구성하였다. 이 타이틀 바는 템플릿 정합률, 입술움직임 임계값, 온라인 템플릿 검출(capture)을 시도한 횟수를 포함한다. 영상 제어부분은 영상프레임률 동적제어, 칼라/2진영상 변환, 정지/동영상 저장, 일시정지/재실행 등을 포함한다. 특히, 템플릿 초기영상을 온라인 하에서 추출한 템플릿으로 대체하기 위한 버튼(TMPL_onck)이

마련되어 있다. 재생부분에서는 영상 제어부분에서 저장한 정지/동영상을 재생해 볼 수 있다. 여기에서, 각 파라미터와 버튼 이름은 편의상 약칭을 정의하여 사용하였다.

다음으로, 온라인 템플릿 대체버튼(TMPL_onck) 설정에 의한 조명환경 적용에 대해 기술하면, 온라인 템플릿 자동 추출 요건이 충족되지 않더라도 온라인 템플릿을 대체할 수 있는 버튼을 설정하였다. 이 버튼을 클릭하면 템플릿 적합도가 현재까지의 최대적합도(CrrTL)에 도달하는 순간, 템플릿을 검출하여 템플릿 초기모델을 대체할 수 있다. 이를 통하여 조명환경에 구애받지 않고 온라인 템플릿 교체가 가능하도록 하였다. 조명환경이 열악한 경우에는 온라인 템플릿 자동추출 요건이 충족되지 않거나 상당한 시간이 지연될 수 있기 때문에 이를 보완한 것이다.

3.6 음성/영상시스템의 연동

우선, 데이터 공유 및 연동 실행에 대해 기술하면, 영상처리와 음성인식은 각각 별개의 윈도우로 나뉘어서 독립적으로 실행된다. 따라서 영상처리와 음성인식 프로세스 사이의 데이터 공유를 위해 IPC(Inter Process Communication) 통신이 요구된다. 이를 위해 공유메모리 방식을 통해 데이터를 공유하도록 하였다[5],[9]. 입술움직임 영상신호는 음성인식 과정에서 확인여부에 구애됨이 없이 공유메모리에 기록만 하는 것으로, 음성인식의 음성구간 검출모듈에 대해 독립적으로 공유메모리를 확인하도록 하였다.

연동기능은 일단 영상처리와 음성인식이 모두 실행되는 상황에서 수행된다. 영상처리부에서는 입술움직임 영상신호가 있는 경우 이를 공유메모리에 기록하고, 음성인식기는 음성구간을 검출하면서 영상신호가 있는지를 확인한다.

다음으로, 음성인식의 음성구간 검출기능 보완에 대해 기술하면, 그림 2. 음성/영상 결합시스템의 상세구조도 내의 음성구간 검출 모듈 ④에 영상 신호값을 읽어들이 수 있는 기능을 부가하였다. 이는 음향에너지 분석 외에 입술 움직임이 있는지의 여부를 추가로 확인하도록 한 것이다. 또한, 영상처리가 처음부터 실행되지 않는 경우 혹은 영상처리 실행이 중도에 멈추는 경우 음성인식 기능은 기존의 방식대로 정상적으로 동작될 수 있도록 구현하였다.

음성구간 검출모듈에서 영상신호 확인을 위한 추가기능의 구현에 있어서 프레임별 처리 시간차에 대한 고려가 요구된다. 즉, 음성처리의 경우 보통 10msec 단위로 음성 프레임이 처리되는데 비해 영상처리의 경우 30msec 단위이하로 영상프레임이 처리되기 때문이다. 따라서 영상검출 과정에서 영상유실이 음성인식에 미치는 영향을 최소화하기 위한 보상차원의 기능을 추가하였다. 이를 위해 영상검출에 성공한 프레임의 데이터가 다음 프레임까지 지속될 수 있도록 조정하였다. 일단 입술움직임 영상이 검출되면 그 결과가 일정 횟수의 프레임까지 지속되도록 하였다.

IV. 실험결과

영상신호 추출과정에서는 입술움직임 영상을 정확히 찾아내는 것이 무엇보다도 중요하다. 영상처리를 실행하면서 추출할 수 있는 입술움직임 영상검출의 결과, 적합도, 임계값 등의 데이

터를 시각적으로 실시간 확인해 볼 수 있도록 하였다. 또한, 영상과 음성인식 연동실험 환경을 구축하여 입술움직임 영상신호를 고려한 음성인식이 실행되는 지를 확인하였다. 실험도구로서 펜티엄(Pentium) IV 3.6GHz의 성능을 가진 PC외에 추가로 장착된 320×240 해상도와 초당 15 프레임의 속도를 유지하는 PC용 칼라 영상카메라가 사용되었다.

우선, 입술움직임 영상검출 결과에 대해 언급하면, 일상의 조명환경 하에서 얼굴 전체의 움직임이 거의 없는 상태에서 입술 움직임을 어느 정도 정확하게 검출하는지를 확인하였다. 이 때, 화자와 카메라 간의 측정간격은 50cm로 유실을 포함한 연동오류가 발생할 때까지 횟수를 측정하는 방식으로 진행되었다. 그 결과, 프로그램 구동으로부터 입술움직임 영상의 초기검출까지의 시간은 무시할 정도였고, 얼굴요소 중에서 가장 많은 검출오류 요소는 눈의 움직임 이었으며, 입술움직임 영상검출의 성공률은 표 3에 나타난 바와 같이 95%에 도달하였다. 그리고, 3.6절에 기술한 바와 같이 입술움직임 검출결과의 지속기능을 추가함으로써 검출오류가 음성/영상 연동에까지 미칠 수 있는 영향을 최소화 하였다.

표 3. 입술움직임 영상검출의 성공률

Table 3. Success rate of lip movement image detection

입술움직임 횟수 (처음 오인식까지)	오류 내역			성공률 (%)
	유실	연동오류	계	
43	1	1	2	95
46	1	1	2	96
39	1	1	2	95
18	-	1	1	94
41	2	1	3	93
61	2	1	3	95
248	7	6	13	95

여기서, 유실이란 입술움직임이 있었는데도 검출하지 못한 경우이고, 연동오류란 입술움직임이 아닌 것 이외의 기타 턱이나 눈의 움직임을 검출하는 경우를 의미한다. 움직임이 없는 경우 어떤 형태로서 움직임을 나타내는 경우는 발견되지 않았다.

표 4는 조명환경, 화자와 영상카메라 간격, 오인식 다발 얼굴요소, 검출 오류 보상기능 등 입술 움직임 영상검출을 위한 실험환경을 나타낸 것이다.

표 4. 입술움직임 영상검출 실험환경

Table 4. Experimental environments of lip movement image detection

조명환경	일반가정/사무실 환경 (300~500Lux, 100Lux 이상가능)
화자와 영상카메라 간격	약 50cm
초기검출시작 소요시간	약 3초
오인식 다발 얼굴요소	눈, 얼굴/배경 경계면
검출 오류 보상기능	명암값 보상, 2단계 검출

다음으로, 영상과 음성인식 연동실험 결과에 대해 기술하면, 영상과 음성을 연동시킨 실험환경 하에서 영상신호의 유무

에 따른 음성인식 결과의 출력여부를 확인하였다. 실험은 오류가 발생할 때까지의 결과를 확인한 것이다. 또한 영상기능을 실행하지 않는 경우와 증도에서 멈추는 경우에도 음성인식 기능이 정상적으로 수행됨을 확인하였다.

표 5는 음성/영상의 연동실험 결과를 나타낸 것이다. 이 표 5에서 입력이 없는 오류(Not input error)는 발생과 함께 입술움직임이 있었는데도 입술움직임 영상신호를 확인하지 못하여 음성인식 진행을 하지 않은 경우이다. 입력이 있는 오류(Input error)는 카메라로 하여금 입술움직임을 검출하지 못하도록 한 상태에서 발생한 경우로 음성인식이 진행된 경우는 없었다. 즉, 음성/영상의 연동결과 조명환경 등 다양한 실험환경하에서 그 연동률을 99.3%까지 높일 수 있어, 음향잡음에 의한 음성인식 실행을 대부분 억제할 수 있었다.

표 5. 음성/영상 연동 실험결과

Table 5. Experimental results of the linked speech and image

발성횟수	오류 내역			성공률(%)
	입력없음	입력있음	계	
72	1	0	1	98.6
103	1	0	1	99.0
193	1	0	1	99.5
199	1	0	1	99.5
567	4	0	4	99.3

그 외 실험으로, 음성/영상 시스템 연동에 앞서, 음성인식 대상이 아닌 음향잡음에 의해서 음성인식 과정이 진행되는 것을 확인하였다. 또한 조명환경의 적용여부를 확인하기 위하여 전등의 일부 혹은 전부를 소등하는 등 조명기기의 밝기를 달리하는 실험을 하였다. 그리고 음성/영상 연동 실험에 의해 순수 음향잡음이 음성인식 과정에서 차단되는지의 여부를 확인하였다. 이들 실험결과를 요약하여 나타내면 표 6과 같다. 이 표에서 음향잡음원은 크기와 종류에 따라 다소 차이가 있을 수 있으나 음향이 클 때 음성인식진행률이 큰 경우 음성으로 오인식되고 있음을 의미한다.

표6. 실험결과 요약

Table 6. Summary of experimental results

음성인식기 잡음대처 실험		
음향잡음원	잡음크기(dB)	음성인식진행(%)
TV 대화	65	95.7
	45	70.6
MP3 음	85	80
	70	50
영상처리 실험 및 조명환경 적용		
추적대상	조명크기(lx)	성공률(%)
입술움직임	사무실(500~300) 가 정(300~100) 실험실(100~)	약 95
음성/영상 연동실험		
입력대상	내용	성공률(%)
음향잡음	음성인식 차단	100
	음성인식 실행	99.3
음성발성	음성인식 미실행	0.7

V. 결론

본 논문에서는 음성인식을 위한 음성구간 검출과정에서 유입될 수 있는 동적인 음향에너지 이외에 화자의 입술움직임 영상신호까지 포함함으로써 외부 음향잡음이 음성인식 대상으로 오인식되는 것을 방지하기 위한 한 가지 방법이 제시되었다. 우선, 연속적인 영상이 PC용 영상카메라를 통하여 획득되고 그 입술움직임 여부가 식별된다. 다음으로, 입술움직임 영상신호 데이터는 공유메모리에 저장되어 음성인식 프로세서와 공유한다. 한편, 음성인식의 전처리 단계인 음성구간 검출과정에서 공유메모리에 저장되어진 데이터를 포함함으로써 화자의 발성에 의한 음향에너지인지의 여부가 입증된다. 최종적으로, 음성인식기와 영상처리를 연동시켜 실험한 결과, 영상카메라에 대면해서 발생하면 음성인식 결과의 출력에 이르기까지 연동처리가 정상적으로 진행됨을 확인하였고, 영상카메라에 대면치 않고 발생하면 연동처리시스템이 그 음성인식 결과를 출력치 못함을 확인하였다. 이러한 결과는 음향에너지가 입력되더라도 입술움직임 영상이 확인되지 않으면 음향잡음으로서 간주하도록 한 것에 따른 것이다. 따라서 다양한 환경에 적응하기 위해 입술움직임 영역을 정확히 추적하는 것도 중요하다. 이를 위해 온라인 하에서 입술움직임 표준 영상의 특징값을 추출하였고, 이를 매개로 하여 오프라인하의 입술움직임 초기 특징값 및 템플릿 초기영상을 온라인하에서 추출된 입술움직임 초기특징값 및 템플릿 영상으로 대체함으로써 입술움직임 영상추적의 변별력을 향상시켰다. 또한, 입술움직임 영상추적과정을 시각적으로 확인하고 실시간으로 관련된 패러미터를 해석하기위해 영상처리 테스트베드를 구축하였다. 음성과 영상처리 시스템의 연동결과 다양한 조명환경하에서도 약 99.3%의 연동율을 나타냈다. 이러한 연동율은 현재의 음성인식 성능을 유지하면서 음향잡음을 차단할 수 있다는 가능성을 나타낸 것이다.

향후의 과제로서, 현재의 PTT (Push to Talk) 음성인식 모드가 Non-PTT 모드로의 전환에 활용될 수 있는지에 관한 검토와 음향잡음을 보다더 억제하기 위한 방법 및 다양한 영상획득 환경에서 강인한 입술움직임 영상검출의 실현 등을 들 수 있다.

참고 문헌

- [1] Lawrence Rabiner, Biing-Hwang Juang, Fundamentals of Speech Recognition, Prentice Hall, pp.11-68, 1993.
- [2] G. Potamianos, & C. Neti, G. Gravier, A. Grag, & A.W. Senior, "Recent advances in the automatic recognition of Audio-visual speech", Proceedings of the IEEE, Vol.91, No.9, pp.1306-1326, 2003.
- [3] Shogo Nishida, "Speech Recognition Enhancement by Lip-Information", Media Laboratory, MIT Cambridge, MA 02139, pp.198-204, April 1986.
- [4] M.T. 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.

[5] G. Potamianos, 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.

[6] S. Nakamura, and E. Yamamoto, "Speech-to-lip movement synthesis by maximizing audio-visual joint probability based on the em algorithm", Journal of VLSI Signal Processing, Vol.27, No.1-2, pp.119-126, 2001.

[7] P. Lucey, & G. Potamianos, "Lipreading using profile versus frontal views", In Proceedings of the International Workshop on Multimedia Signal Processing, (Victoria, Canada), pp.24-28, 2006.

[8] A. W. Liew, S. H. Leung, and W. H. Lau, "Lip contour extraction from color images using a deformable model", Pattern Recognition, Vol.35, No.12, pp.2949-2962, 2002.

[9] G. Potamianos, & C. Neti, "Audio-visual speech recognition in challenging environments", In Proceedings of the European Conference on Speech Communication and Technology, (Geneva, Switzerland), pp.1293-1296, 2003.

[10] A. Liew and S. Wang, "Visual Speech Recognition: Lip Segmentation and Mapping", editors, IGI Global, 2009.

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

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

[13] V. Libal, J. Connell, G. Potamianos, and E. Marcheret, "An embedded system for in-vehicle visual speech activity detection", In proceedings of the International Workshop on Multimedia Signal Processing(MMSP 2007), pp.255-258, Chania, Greece, 2007.

[14] 김응규, 이수중, "입술움직임 영상신호를 활용한 음성 구간 검출", 2007년 한국신호처리시스템학회 추계학술대회 논문집, 제8권, 제2호, pp.187-192, 2007.

[15] 김응규, 최정훈, 이수중, "연속영상 프레임으로부터 입술움직임 영상의 검출방법", 2008년 한국신호처리시스템학회 추계학술대회 논문집, 제9권, 제2호, pp.433-437, 2008.

[16] 김응규, 최정훈, "영상 환경에 적응하는 강인한 입술움

직임 영역 추적법", 2009년 한국신호처리시스템학회 추계학술대회 논문집, 제10권, 제1호, pp.77-80, 2009.

[17] G. Potamianos, C. Neti, J. Luettin and I. Matthews, Audio-visual automatic speech recognition: An overview, in issue in Visual Speech Processing, MIT Press, 2004.

[18] F. Leymarie and M.D. Levine, "Simulating the Grassfire Transform Using the Active Contour Model", Trans. IEEE Pattern Analysis and Machine Intelligence, Vol.14, No.1, pp.56-75, 1992.



김응규(Eung-kyeu Kim)

1976년 02월 충남대학교 공업교육과(학사)
 1978년 02월 충남대학교 대학원 공업교육과(석사)
 1993년 09월 오사카대학 대학원 정보공학과(박사)
 1987년 10월 ~ 1989년 09월 교오대학 공학부 연구원

1990년 02월 ~ 1992년 03월 오사카대학 기초공학부
 문부교관(助手)

1994년 03월 ~ 현재 한밭대학교 정보통신공학전공 교수
 2003년 05월 ~ 2006년 02월 한밭대학교 BK21지역대학
 육성사업단장

2005년 03월 ~ 2007년 02월 한밭대학교 정보통신전문대
 학원장

2009년 01월 ~ 현재 대한전자공학회 대전충남지부장

2009년 08월 ~ 2010년 07월 오클라호마주립대학교
 컴퓨터과학과 연구교수

※ 주관심분야 : 영신호처리, 컴퓨터비전, 영상압축통신
 컴퓨터그래픽(3D)