

논문-00-5-1-12

블럭 정합을 이용한 비디오 자막 영역의 원 영상 복원 방법

전병태*, 이재연*, 배영래*

A Method for Reconstructing Original Images for Captions Areas in Videos Using Block Matching Algorithm

Byung Tae Chun*, Jae Yeon Lee*, and Younglae Bae*

요 약

이미 방송된 비디오 영상으로부터 자막 영역을 제거하고 원 영상으로 복원할 필요가 종종 발생한다. 복원될 영상의 양이 적을 경우 수 작업에 의한 복원이 가능하나, 비디오 영상과 같이 복원할 영상이 많아질 경우에는 수 작업에 복원은 어렵다고 볼 수 있다. 따라서 자동으로 자막 영역을 원 영상으로 복원할 수 있는 방법이 필요하게 된다.

기존의 영상 복원에 관한 연구는 주로 블러링(blurring)된 영상을 주파수 필터를 사용하여 선명하게 복원하거나, 영상 통신을 위한 비디오 코딩 방법에 대한 연구가 많이 이루어졌다.

본 논문에서는 블럭 정합 알고리즘(Block Matching Algorithm)을 이용하여 자막 영역을 복원하는 방법을 제안하고자한다. 자막 복원을 위한 사전 정보로 자막 영역 정보와 장면 전환 정보를 추출한다. 추출된 자막 정보로부터 자막의 시작 프레임, 끝 프레임, 자막 문자의 구성 요소 정보를 얻을 수 있다. 자막 정보(자막의 시작 프레임, 끝 프레임)와 장면 전환 정보를 이용하여 복원의 방향성 및 복원의 중점을 결정한다. 복원의 방향성에 따라 각 프레임마다 문자의 구성 요소에 대한 블럭 정합을 수행하여 원 영상을 복원한다. 실험결과 비교적 움직임이 적은 영상에서는 복원이 잘 됨을 볼 수 있었으며, 복잡한 배경을 갖고 있는 영상의 경우도 복원됨을 볼 수 있었다.

Abstract

It is sometimes necessary to remove the captions and recover original images from video images already broadcast. When the number of images requiring such recovery is small, manual processing is possible, but as the number grows it would be very difficult to do it manually. Therefore, a method for recovering original image for the caption areas is needed. Traditional research on image restoration has focused on restoring blurred images to sharp images using frequency filtering or video coding for transferring video images.

This paper proposes a method for automatically recovering original image using BMA(Block Matching Algorithm). We extract information on caption regions and scene change that is used as a prior-knowledge for recovering original image. From the result of caption information detection, we know the start and end frames of captions in video and the character areas in the caption regions. The direction for the recovery is decided using information on the scene change and caption region(the start and end frame for captions). According to the direction, we recover the original image by performing block matching for character components in extracted caption region. Experimental results show that the case of stationary images with little camera or object motion is well recovered. We see that the case of images with motion in complex background is also recovered.

* 한국전자통신연구원/ 컴퓨터 소프트웨어 기술연구소
Electronics and Telecommunications Research Institute/CSTL

I. 서 론

최근 컴퓨터 하드웨어·소프트웨어 기술의 급속한 발전

으로 자료 저장의 대용량화, 처리 속도의 고속화, 통신에 의한 정보의 공유화 등 많은 사회적, 문화적, 기술적 변화가 일어나고 있다. 이러한 컴퓨터 기술의 발전은 여러 기술 분야의 통합화 추세로 발전하고 있으며 그 중 하나가 동영상(또는 비디오 영상) 처리 기술 분야이다.

멀티미디어 기술은 다 매체(영상, 음성, 음향, 등)를 효율적으로 처리하여 사용자에게 이용하기 편리한 최적의 형태를 제공하는 기술이라고 볼 수 있다. 멀티미디어 기술에 사용되는 다 매체 중 동영상 매체의 사용은 날로 증가하는 추세이며, 현재 동영상 처리 및 관리에 관한 연구가 활발히 진행되고 있다.

비디오 매체는 방송, 교육, 가정, 영화 등과 같이 많은 분야에 사용되어 지고 있다. 비디오 자막은 사용자에게 비디오 내용에 관한 이해를 돕기 위하여 문자화된 형태로 비디오에 삽입된다. 삽입된 비디오 자막은 비디오에 관한 의미 정보를 함축적으로 나타내거나 비디오 내용에 대한 의미 전달 수단으로 사용되고 있다. 예를들어, 뉴스 비디오의 뉴스 아이콘 자막은 앞으로 설명 될 뉴스 기사에 대한 축약된 형태의 의미 정보를 제공한다. 외국에서 제작된 영화 방송물(영화, 다큐멘터리, 뉴스 등)을 수입하여 자국에서 방송할 경우, 영상에 자국어 자막을 삽입 해줌으로써 방송물에 대한 의미 전달에 자막이 중요한 수단으로 사용된다.

복원의 필요성은 다음과 같은 3가지 이유에서 필요하다고 볼 수 있다. 첫째, 자막 복원은 복원하고자하는 영상의 량이 적을 경우 수 작업 처리가 가능하나 동영상의 경우처럼 처리하여야할 영상의 량이 많을 경우 수 작업에 의한 복원은 처리 시간 많이 소요됨으로써 복원이 불가능할 수 있다. 따라서 자동으로 자막 영역을 원 영상으로 복원하는 방법이 필요하게 된다. 둘째, 영상의 재사용이라는 관점에서 반드시 필요한 기술이라고 볼 수 있다. 방송국에서 자막이 삽입된 영상을 필요에 따라 원 영상으로 복원하여 재 사용할 필요성이 발생할 경우, 자막 영역을 원 영상으로 복원하여 방송함으로써 시청자에게 시청의 편의성을 제공 할 수 있다. 셋째, 외국에서 제작된 영상에 외국어 자막이 삽입되었을 경우, 한 국가에서 지역적으로 사용되는 언어가 다를 경우, 방송물에 삽입된 자막을 제거하고 원 영상으로 복원한 후 원하는 언어(또는 자막)로 대체함으로써 다른 자막으로 대체의 편리성을 제공 할 수 있다.

기존의 영상 복원에 관한 연구는 주로 블러링(blurring)된 영상을 주파수 필터링을 사용하여 원 영상으로 복원^{[1][2]}하거나 압축 영상 전송 시 발생하는 손실을 복원^{[3][4]}하는 연구가 많이 이루어졌다. 이 연구 방법들은 영상을

보다더 선명하게 개선하기 위하여 복원하거나, 통신상에서 발생하는 잡음을 제거함으로써 손실된 영상을 복원하고자 했다. 국보급 문화재 복원 방법^[5]은 보간법(interpolation)에 의한 선 위주의 복원 방법이 연구되어졌다. 보간법에 의한 원영상 복원은 선 위주의 복원 방법이기 때문에 자막 영역처럼 복원 영역이 영역 단위일 경우 부적합하다고 볼 수 있다.

본 논문에서는 비디오 정보(자막 정보, 장면 전환 정보)와 블럭 정합 방법(BMA: Block Matching Algorithm)을 이용하여 자막 영역을 원 영상으로 복원하는 방법을 제안하고자 한다. 제안된 영역 복원 방법에 대한 전체적인 흐름을 살펴보면 그림 1과 같다. 동영상에서 자막 복원을 위한 사전 정보로 장면 전환 정보와 자막 정보를 추출한다. 동영상 장면 전환 정보는 장면이 전환되는 지점에 관한 정보이다. 자막 정보는 자막이 시작하는 프레임과 끝나는 프레임에 관한 정보와 자막 문자의 구성 요소 정보를 말한다. 동영상의 장면 전환 정보와 자막 정보(자막의 시작과 끝 정보)를 이용하여 복원의 방향성을 결정하고, 복원의 방향성에 따라 자막 문자의 구성 요소 단위로 원 영상과 블럭 정합을 수행하여 자막 영역을 원 영상으로 복원한다.

논문의 구성은 다음과 같다. 제1장에서는 기존 연구 방법 및 자막 영역에 대한 원 영상 복원의 필요성에 대하여 살펴보고, 제2장에서는 비디오 자막의 특성과 복원을 위한 사전 정보(장면 전환 정보, 자막 정보) 추출에 대하여 설명하겠다. 제3장에서는 복원의 방향성 결정 방법에 대하여 설명하고, 결정된 복원의 방향성에 따라 자막의 구성 요소 단위로 블럭 매칭을 수행하여 자막 영역을 원 영상으로 복원하는 방법에 대하여 설명하겠다. 제4장에서는 제

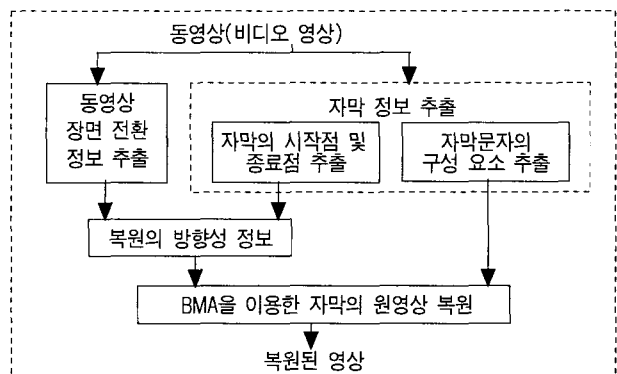


그림 1. 블럭 정합을 이용한 비디오 자막의 원 영상 복원
 Fig. 1. Original image recovery for caption region in videos using BMA

안된 원 영상 복원 방법의 실험 결과를 보여주고 그 결과에 대하여 설명하였다. 제5장에서는 향후 연구 방향에 대하여 설명하였다.

II. 비디오 자막의 특성 및 비디오 정보 추출

1. 비디오에 삽입된 자막의 특성

비디오에 삽입된 자막 영역은 다음과 같은 몇 가지 특성을 갖고 있으며, 본 논문에서는 그 특성을 이용하여 자막 영역을 원 영상으로 복원에 한다.

첫째, 복원할 자막 영역은 실제 자막 문자 영역보다 크다. 비디오에 삽입된 자막 문자는 문자 영역을 강조하기 위하여 그림 2와 같이 문자 주위에 문자의 윤곽선을 삽입한다. 이러한 문자의 윤곽선 영역 역시 원 영상을 손상시켰다고 볼 수 있으므로 원 영상 복원 시 복원에 참여시켜야 한다. 그렇지 않을 경우 문자의 윤곽선 영역 처리를 별도로 처리하여야 하는 문제가 발생한다.

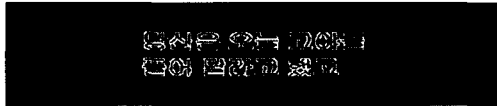


그림 2. 자막의 윤곽선 영역
Fig. 2. Contour line for caption

둘째, 그림 3과 같이 자막이 시작되는 전 프레임이나 자막이 끝나는 다음 프레임은 자막이 삽입되지 않는 원 영상을 갖고 있다. 이러한 원 영상은 복원을 위한 중요한 정보로 사용할 수 있다. 여기서 $St-fr_i$ 는 추출된 문자 영

역의 i 번째 군집의 시작 프레임이고, $Ed-fr_i$ 는 추출된 문자 영역의 i 번째 군집의 종료 프레임이다.

셋째, 비디오 자막은 주로 대화나 설명을 위한 장면에서 주로 삽입된다. 따라서 다른 장면에 비하여 물체(사람 또는 사물)의 움직임 비교적 적고 카메라의 움직임이 비교적 단순함을 볼 수 있다.

넷째, 자막이 삽입된 동영상에서는 장면 전환이 자주 발생하지 않는다. 왜냐하면, 3번째 가정에서 언급했듯이 대부분의 자막 영역은 대화 및 설명 장면이므로 급격한 장면 전환은 자주 발생하지 않는 특성이 있다. 장면 전환이 발생한다면 보통 1~3회 정도 발생하며 3번 이상의 장면 전환은 거의 발생하지 않는다.

2. 비디오 자막 정보 추출

기존의 문자 영역 추출 방법^{[6][7]}은 전체 영상에 대하여 컬러 축소(color reduction), 영역 분할 및 합병(region split and merge), 질감 분석(texture analysis) 등과 같은 방법을 이용하여 문자 영역을 추출했다. 이 방법들은 많은 휴우리스틱(heuristic) 변수와 추출하고자 하는 문자의 사전 지식에 의해 임계치 값을 설정함으로써 알고리즘을 일반화하기 어렵다는 문제점이 있다.

본 논문에서 사용한 자막 추출 방법은 지형학적 특징점(topographical features)을 이용한 자막 영역 추출 방법^[8]을 사용하였다. 이 연구 방법은 추출된 문자의 지형학적 특징점을 점-선-면 확장법을 이용하여 문자 영역을 추출함으로서 기존 문자 영역 추출의 문제점인 휴우리스틱 변수의 사용을 최소화하고 임계치 값을 일반화함으로서 일반화된 문자 영역 추출 방법을 제안하였다. 지형학적 특징

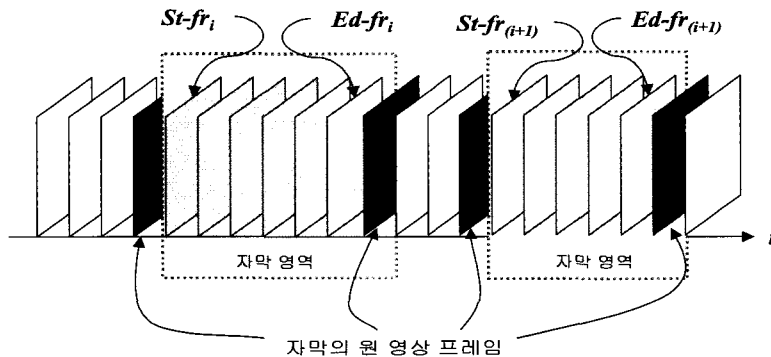
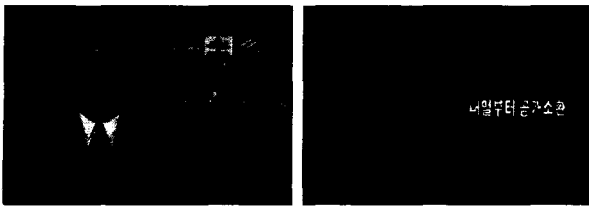


그림 3. 자막 영역의 원 영상 프레임
Fig.3. Original image frame for caption region



(a) 비디오 자막 영상 (b)추출된 자막 영역

그림 4. 지형학적 특징점을 이용한 자막 영역 추출
Fig. 4. Extracting caption regions using topographical features

점을 이용한 자막 영역 추출 결과는 그림 4(b)에서 보여 주고 있다.

자막 추출 결과 복잡한 배경과 자막의 특성으로 자막 영역이 정확히 추출되지 않고 그림 5(a)와 같이 간혹 오류가 발생할 수 있다. 이러한 오류는 연속된 프레임과의 관계를 조사하여 연속된 프레임에서 갑자기 자막 영역이 발견되거나 자막이 없는 경우, 이런 영역은 오류 영역일 가능성이 많으므로 그림 5(b)와 같이 오류 수정을 한다. 예를들어 20개 프레임에서 문자 영역들이 발견되었다고 할 때, 이 영역들을 모두 중첩시켜 영역의 중첩율을 계산한다. 20개 프레임의 영역 중에서 동일 영역에 약 80%(16개 프레임)의 중첩율이 발생하면 문자 영역으로 보고 그렇지 않으면 오류라고 볼 수 있다. 따라서 80% 미만의 중첩율을 보이는 영역은 오류로 간주하고 제거하고, 80% 이상의 중첩율의 영역에서 그 영역이 존재하지 않은 프레임이 존재하면 이 역시 오류이므로 강제적으로 그 프레임에 영역을 만들어 준다.

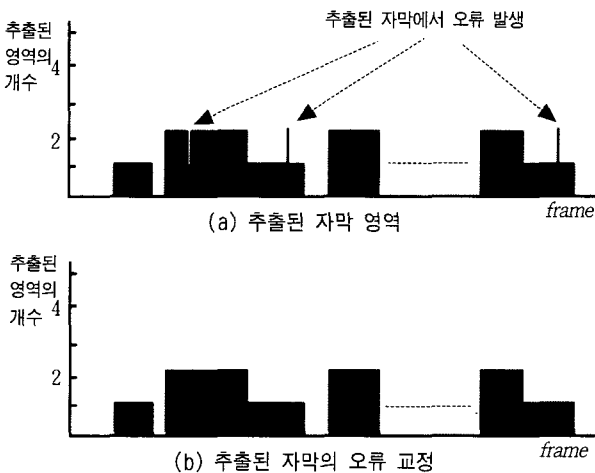


그림 5. 추출된 자막 추출의 오류 수정
Fig. 5. Error correction for extracting caption area

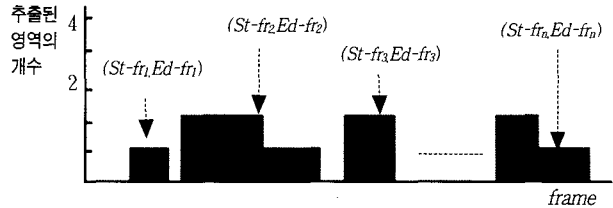


그림 6. 에러 수정 후 자막의 시작점 끝점 추출
Fig. 6. Extracting start and end frame for caption

오류 수정 후, 복원의 방향성 정보 결정을 위하여 그림 6과 같이 추출된 자막의 시작 프레임($St-fr_1$)과 끝 프레임($Ed-fr_1$) 정보를 추출한다.

자막이 영상에서 차지하는 영역이 크기 때문에 선(line)에 기초한 복원 방법이 적합치 않고 영역 단위의 복원 방법이 적합하다고 볼 수 있다. 영역 단위의 복원을 위하여 복원의 최소 단위인 문자의 구성 요소를 구하기 위하여 라벨링(labeling) 연산을 수행한다. 그림 7은 그림 2의 문자 영역을 라벨링 연산을 수행한 결과를 보여 주고 있다.



그림 7. 추출된 문자 영역 및 문자 구성 요소 영역
Fig. 7. Extracted caption area and character components

3. 비디오 장면 전환 정보 추출

자동화된 비디오 내용 분석에 있어서 동영상 장면 자동 분할은 중요한 연구 분야 중에 하나이며, 이 연구에서 중요한 연구 주제는 2가지로 분류 될 수 있다. 첫째, 장면 전환 전환 지점을 정확히 추출하기 위하여 사용될 특징들을 어떻게 추출할 것 인가이다. 둘째, 추출된 특징을 어떤 메카니즘을 이용하여 장면 전환 지점을 찾을 것인가에 관한 것이다. 기존의 연구 방법^{[9][10][11]}은 위의 두 문제를 동시에 해결하기 위한 방법을 모색하려는 연구가 이루어졌다.

본 논문에서 사용한 장면 전환 정보 추출 방법^[11]은 보통의 장면 전환 뿐만아니라 카메라 특수 효과 장면을 자동으로 분할 할 수 있는 방법을 사용하였다. 이 방법에 사용된 특징 선택 방법을 살펴보면 다음과 같다.

장면 분할에 사용될 특징은 차분 영상의 히스토그램 분산, 화소 값의 차이, 히스토그램의 차이, 윤곽화소 수의 차이 등 여러 종류가 있을 수 있으나 특징의 유효성 평가

및 처리 시간을 비교하여 다음과 같은 3가지 특징들을 선택하였다. - 차분 영상의 히스토그램(식(1)), 화소 값의 차이(식(2)), 히스토그램 분산(식(3)).

$$D_h = \frac{\sum_{i=1}^K |H_m(i) - H_{m+1}(i)|}{K} \quad (1)$$

$$D = \frac{\sum_{i=1}^K (H_{m,m+1}(i) - H)^2}{K} \quad (2)$$

$$D_p = \frac{\sum (I_m(x, y) - I_{m+1}(x, y))}{N} \quad (3)$$

식(1)에서 사용된 $H_{m,m+1}$ 은 m -th과 $(m+1)$ -th 프레임 사이에서 계산된 히스토그램 차(difference)이고, K 는 차 값의 범위, H 는 $H_{m,m+1}$ 의 평균이다. 식(2)에서 사용된 $I_m(x, y)$ 는 좌표 (x, y) 에서 m -th 프레임의 화소 크기(pixel intensity)이고, N 은 한 프레임의 총 화소의 개수이다. 식(3)에서 사용된 H_m 은 m -th 프레임의 히스토그램이고, K 는 화소값의 범위이다.

추출된 특징들을 효율적으로 분류하기 위한 메카니즘은 멀티 레벨 임계치 법, 최소거리 법, 최대 유사도법 등 다양한 방법이 제안되어 있으나, 그 중에서도 최대 유사도법이 가장 널리 사용되고 있다. 그러나 이와같은 통계적인 접근 방법은 대체적으로 각 특징들이 정규분포를 따른다는 가정을 하고 있는데 실제로 이와같은 가정은 성립하지 않는 경우가 많다. 이에 반하여 신경망 알고리즘은 그와같은 가정을 두지 않으며, 신경망이 가지는 Non-parametric한 특성 때문에 비교적 적은 양의 학습 정보만으로도 훈련시킬 수 있다는 장점을 가지고 있다. 따라서 장면 전환 지점 추출을 위한 메카니즘으로 오류 역전파(Back-Propagation) 신경망 알고리즘^[12]을 사용하였다.

신경망의 구조는 3가지 특징을 입력하기 위한 3개의 입력층, 3개의 은닉층 그리고 장면 전환을 판단하는 1개의 출력층으로 이루어 졌다. 추출된 3가지 특징들을 학습하기 위하여 특징들은 정규화(Normalization) 후 신경망의 입력층에 입력하게 된다. 신경망을 이용하여 추출된 장면 전환 정보는 복원의 종료점 결정에 이용한다.

III. 블럭 정합을 이용한 자막 영역의 원 영상 복원

1. 원 영상 복원의 방향성 결정

추출된 자막 정보(자막의 시작 및 끝 위치 정보) 및 장면 분할 정보는 원 영상 복원의 출발점과 복원의 방향성을 결정하는 중요한 정보로 이용된다. 복원의 방향성을 결정하는 이유는 자막의 시작되는 전 프레임이나 자막의 끝나는 다음 프레임은 자막이 삽입되지 않은 원 영상을 갖고있기 때문에 이 영상을 원 영상 복원에 중요한 정보로 이용하려는 것이다. 복원의 방향성은 추출된 자막 정보에 따라 결정되며, 어느 방향으로 어디까지 복원을 진행할 것인가를 결정한다.

추출된 장면 전환 정보와 자막 정보를 이용하여 결정된 복원의 방향성과 복원의 종점을 분류하여 보면 그림 8과 같이 3가지로 분류 할 수 있다.

첫째, 그림 8(a)와 같이 자막 영역에 장면 분할이 발생치 않을 경우, 자막 복원의 방향성과 종점은 자막의 시작 프레임부터 중앙 지점까지 순방향으로 복원하고, 끝 프레임에서 중앙 지점까지 역 방향으로 복원을 한다. 종점 Mid_fr_{12} 와 Mid_fr_{21} 은 동일한 값을 갖게되며 자막의 중앙 지점의 값을 갖게 된다.

둘째, 그림 8(b)와 같이 장면 분할이 1회 발생 할 경우,

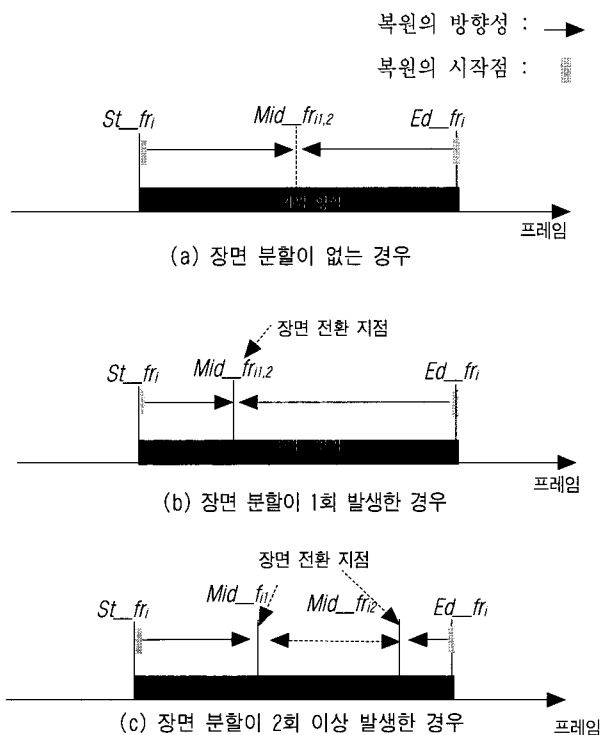


그림 8. 복원을 위한 방향성과 시작점 결정
Fig. 8. Deciding the direction for recovery and start point

시작 프레임에서 장면 분할 지점까지, 끝 프레임에서 장면 분할 지점까지 복원한다. 중점 Mid_fr_{11} 와 Mid_fr_{12} 은 장면 분할이 발생된 지점의 값을 갖게된다.

셋째, 그림 8(c)와 같이 장면 분할이 2회 이상 발생하는 경우, 자막의 시작 프레임부터 장면 분할 지점까지, 자막의 끝 프레임부터 장면 전환 지점까지 복원한다. 중점 St_fr_i 시작하여 Mid_fr_{11} 까지 되며, Ed_fr_i 에서 시작하여 역 방향으로 Mid_fr_{12} 까지 된다.

그러면 장면 분할이 2회 이상 발생된 지점을 어떻게 처리할 것인가? 본 논문에서는 이 부분에 대한 처리는 고려하지 않았다. 왜냐하면, 장면 분할 지점과 지점 사이에는 원 영상 정보가 존재하지 않기 때문에 참조할 원 영상이 없게된다. 따라서 본 방법으로는 복원이 어렵다고 볼 수 있다. 이런 문제를 해결하기 위하여 기존에 사용한 보간법(interpolation)을 사용하거나 다른 처리 방법의 사용 모색하여야 된다. 대부분 자막은 2.1절의 설명한 자막의 특성에서 볼 수 있듯이 대부분 자막이 존재하는 곳은 대화나 설명 장면이므로 급격한 장면 전환이 2회 이상 발생하는 경우는 거의 없다고 볼 수 있다.

2. 블럭 정합을 이용한 자막 영역의 원 영상 복원

자막 영역을 복원하기 위해서 복원 하고자 하는 자막 영역이 원 영상의 어느 부분과 밀접한 관계가 있는지를 찾아내야 한다. 원 영상 영역을 찾기 위하여 블럭 정합 방법^[13]을 사용한다. 블럭 정합 방법은 동영상 압축 및 전송에 많이 사용되고 있는 방법으로 현재 프레임과 참조 프레임에서 최대 유사도를 찾기 위한 방법으로 사용된다. 블럭 정합의 기준은 최소 절대차(MAD : Minimum

Absolute Difference)이며 식(4)와 같다.

$$MAD(d_1, d_2) = \sum_{(x,y) \in R} |F(x,y,t) - F(x+d_2, y+d_1, t-1)| \tag{4}$$

여기서 R 은 문자 구성요소 단위의 문자의 경계 화소, $|d1|, |d2|$ 값은 탐색 범위 보다 작은 값이고 $F(x,y,t)$ 는 시간 t 프레임에 있는 좌표 (x,y) 의 화소를 말한다.

블럭 정합 방법에 의한 자막 영역의 원 영상 복원 방법은 그림 9에서 보여지고 있다. 그림 9(b)는 복원될 자막 문자의 영역을 나타내고 있다. 비교될 자막 영역은 문자 단위가 아니라 2.1절에서 구해놓은 문자의 구성 요소 단위이다. 또한 문자의 구성 요소 영역 전체에 대하여 블럭 정합을 수행하는 것이 아니라 문자의 윤곽선을 포함하는 외곽선 화소에 대하여 블럭 매칭을 수행한다.

그림 9(a)는 복원될 자막 영역의 이전 시점에 있는 참조될 원 영상 프레임이다. 원 영상과 복원될 문자의 구성 요소 외곽선 화소와 탐색 범위(d_1, d_2) 안에서 블럭 정합을 수행하여 최소 절대 차를 만족하는 위치를 찾는다. 최소 위치가 찾아지면 문자의 구성 요소 영역에 원 영상을 대치 시켜 자막 영역을 원 영상으로 복원시킨다. 본 논문에서 사용한 탐색 범위는 $16*16$ 이다.

복원된 영상은 다음 프레임에 있는 자막 영역의 복원을 위하여 원 영상으로 사용된다.

블럭 정합 방법을 이용한 자막 영역 원 영상 복원 방법을 알고리즘화 하면 그림 10과 같다. 알고리즘은 크게 2 단계로 이루어 졌다. 복원을 위한 사전 정보 추출 단계와 추출된 정보를 이용하여 자막 영역을 원 영상으로 복원하는 단계로 구성된다.

사전 정보 추출 단계는 자막영역을 추출하는 *Extracting*

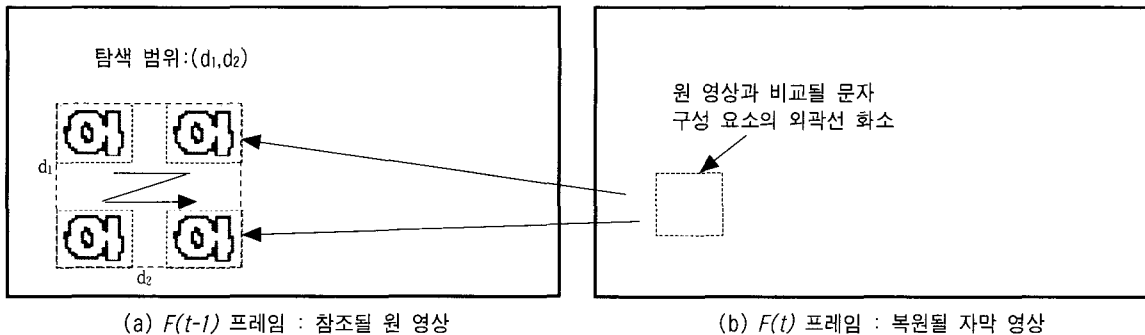


그림 9. 블럭 정합 방법을 이용한 원 영상 복원
Fig. 9. Recovering original using BMA

Pre-Scane : Extracting Caption and Video Information

```

Extracting_Caption_Region(Frames, St_fr, Ed_fr, Caption_Area):
Error_Correction_Extracted_Caption_Region(Frames,St_fr, Ed_fr, Caption_Area):
Extracting_Caption_Component(Frames,St_fr, Ed_fr,Caption_Area,Caption_Compt):
Extracting_Shot_Boundary_Information(Frames,St_fr,Ed_fr,Mid_fr1, Mid_fr2):
    
```

Recovering original image using Block Matching Algorithm

```

while(i != n) // n is a total group number of caption
{
    for(St_inx = St_fr; St_inx < Mid_fr1; St_inx++)
        Call_Recovery_Function(Pre_Frame[St_inx-1],Curr_Frame[St_inx],Caption_Compt[i]):

    for(Ed_inx = Ed_fr; Ed_inx >= Mid_fr2; Ed_inx--)
        Call_Recovery_Function(Pre_Frame[Ed_inx-1],Curr_Frame[Ed_inx],Caption_Compt[i]):

    i++;
}

Call_Recovery_Function(Pre_Frame, Curr_Frame, Caption_Compt)
{
    while(Caption_Compt != End)
    {
        Min_MAD = 999999;
        for(i=Caption_Compt_x-8; i<Caption_Compt_x+8; i++)
            for(j=Caption_Compt_y-8; j<Caption_Compt_y+8; j++)
            {
                MAD = Block_Matching_Operation(Pre_Frame,Curr_Frame,Caption_Compt,i,j):
                if(MAD < Min_MAD)
                {
                    Find_x = i;
                    Find_y = j;
                    MAD = Min_MAD:
                }
            }
        Recovering_Original_Image(Pre_Frame,Curr_Frame,Caption_Compt,Find_x,Find_y):
    }
    Pre_Frame = Curr_Frame:
}
    
```

그림 10. 자막 영역의 원 영상 복원을 위한 알고리즘

Fig. 10. Algorithm for recovering original image for caption areas

Caption_Region() 함수, 추출된 자막 영역의 오류를 수정하는 *Error_Correction_Extracted_Caption_Region()* 함수, 자막 문자의 구성 요소를 추출하는 *Extracting_Caption_Component()* 함수 그리고 장면 전환 정보를 검출하는 *Extracting_Shot_Boundary_Information()* 함수로 구성된다. 여기서 사용된 변수를 살펴보면 다음과 같다. *frame*은 입력 비디오 영상이며 *St_fr*, *Ed_fr*는 *i* 번째 군집 자막의 시작 프레임과 끝 프레임을 나타낸다. *Caption_Area*은 *i* 번째 군집 자막에 소속된 자막 영역을

나타낸다. *Caption_Compt*는 자막 영역에서 추출된 자막 문자의 구성 요소를 나타낸다. *Mid_fr1,2*는 *St_fr*, *Ed_fr* 범위에서 발생하는 장면 전환 지점을 나타낸다.

자막 영역을 원 영상으로 복원하는 단계는 *Call_Recovery_Function()* 함수를 이용하여 자막을 복원한다. 원 영상을 갖고있는 이전 프레임(*Pre_Frame*)과 자막 영역이 있는 현재 프레임(*Curr_Frame*)과의 관계를 블록 정합 방법을 이용하여 자막 영역을 복원한다. 복원은 문자의 구성요소(*Caption_Compt*) 단위로 복원한다.

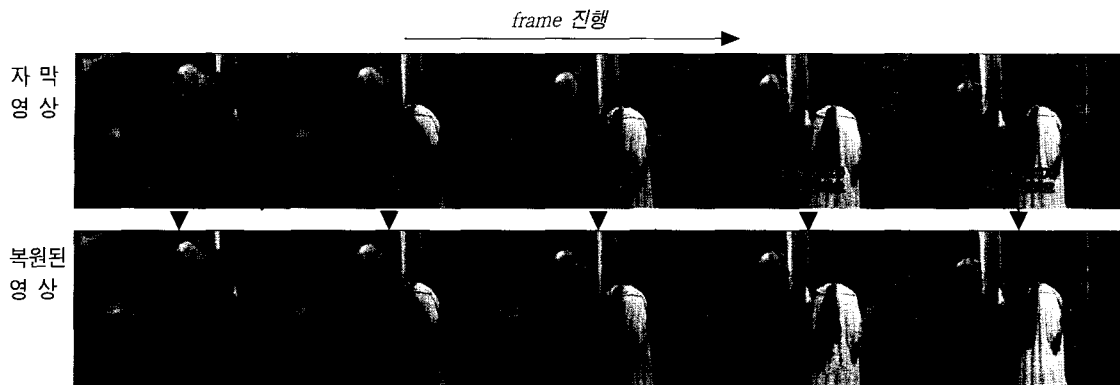


그림 11. 영화 "마야"의 원영상 복원
Fig. 11. Original recovery for a movie "maya"

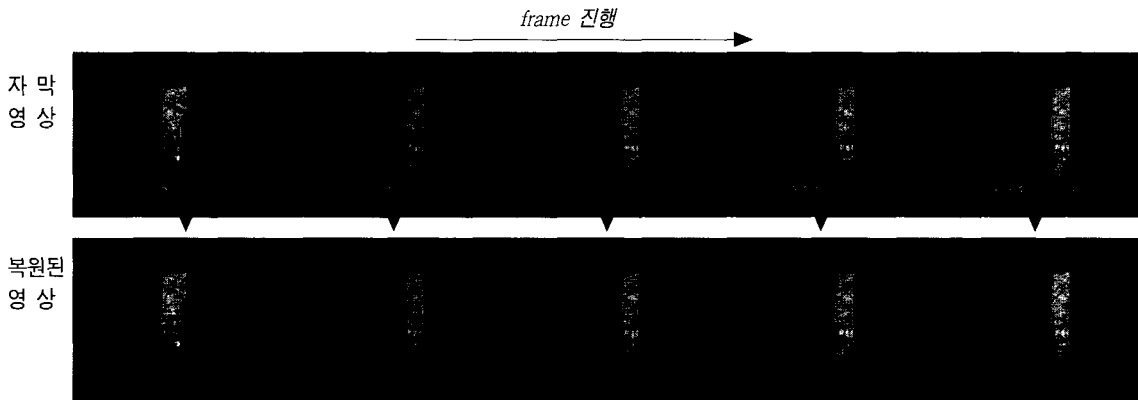


그림 12. 영화에 대한 원 영상 복원
Fig. 12. Original recovery for a movie

IV. 실험 및 결과

실험 환경은 Windows NT상에서 450 MHz 속도를 갖는 Silicon Graphics 320[™] Visual workstation으로 구현하였다. Visual C++6.0 언어를 사용하였으며 비 압축 비디오 영상을 실험 영상으로 사용하였다. 비디오 영상은 RT5 Encoding 보드를 사용하여 file.mpg 파일을 획득한 후, 소프트웨어적으로 Decoding하여 실험에 사용하였다.

실험에 사용된 영상은 영화를 실험 대상으로 실험하였다. 제2.1절의 비디오 자막 특성에서 가정했듯이 자막이 삽입된 곳은 대부분 대화 장면이나 설명을 위한 장면으로 물체의 움직임이나 카메라 모션이 크게 발생하지 않는 것을 볼 수 있었다.

실험결과 물체나 카메라의 움직임이 비교적 적은 영

상에서는 그림 11과 같이 자막 영역이 원 영상으로 완벽히 복원되는 것을 볼 수 있었다. 그림 11은 주인공과 대화하는 장면이며, 물체(여자 배우)가 왼쪽에서 오른쪽으로 이동하고 이동 물체에 자막이 삽입된 것을 볼 수 있다.

그림 12는 일본어 자막이 삽입된 영화에 대하여 원 영상 복원 결과를 보여주고 있다. 배우가 방안으로 들어가는 장면이며 배우의 움직임에 대하여도 블럭 정합을 이용한 자막 영역 복원이 잘 됨을 볼 수 있다.

SF 영화나 액션 영화와 같이 급격한 카메라 동작이나 물체의 이동이 발생하는 자막의 경우, 복원 후 결과가 부분적으로 왜곡 현상이 발생됨을 볼 수 있었다. 이런 문제를 해결하기 위하여 물체의 이동 정보 또는 카메라 움직임 정보를 이용하여 자막 복원을 수행하여야 된다고 생각한다.

IIV. 결 론

실험 결과 정지 영상 또는 움직임이 적은 영상의 경우 복원이 잘 됨을 볼 수 있었다. 그러나 복잡한 배경에 움직임이 다양한 영상의 경우 약간의 왜곡 현상을 볼 수 있었다.

이런 문제를 개선하기 위하여 향후에는 다음과 같은 연구를 수행하여 보완하려 한다. 첫째, 복원의 정확성을 향상시키기 위하여 장면 분할 기술과 자막 추출 기술을 보완하고자 한다. 둘째, 자막 영역에서 장면 분할이 2회 이상 발생하였을 경우 파노라마 기법을 이용하여 복원하는 방법을 연구하고자 한다.

참 고 문 헌

- [1] H. C. Andrews and B. R. Hunt, "Digital Image Restoration," *Prentice Hall, New York*, 1997.
- [2] S. C. Park, M. G. Kang, "Edge-Preserving Image Restoration Algorithm," *International Technical Conference ITC-CSCC'2000*, vol. 1, pp. 23-26, Jul. 1998. Korea.
- [3] ITU-T Recommendation H. 263, Video Coding for Low Bit Rate Communication, Mar. 1996.
- [4] H. Reeve and J. S. Lim, "Reduction of Blocking Effects in Image Coding," *Optical Engineering*, vol. 23, no. 1, pp. 34-37, Jan. 1984.
- [5] 오원근, 유병문, 이인동 외 2인, "영상 처리 기술을 이용한 조사당 벽화의 복원", *한국과학기술연구원/시스템공학연구소 연구보고서(E21131)*, 1992.
- [6] Shoji kurakake Hidetaka Kuwano Kazumi Odaka, "Recognition and Visual Matching of Text Region in Video for Conceptual Indexing," *SPIE Image and Video Processing IV*, vol. 2666, pp. 180-188, Feb. 1996.
- [7] Rainer Lienhart and Frank Stuber, "Automatic Text Recognition in Videos," *SPIE Storage and Retrieval for Image and Video DB V*, Vol. 3022, pp368-378, Feb. 1997.
- [8] 전병태, 배영래, 김태윤, "일반화된 문자 및 비디오 자막 영역 추출 방법", *정보과학회논문지: 소프트웨어 및 응용*, vol. 27, no. 6, pp. 632~641, 6. 2000.
- [9] J. S. Boreczky and L. A. Rowe, "Comparison of Video Shot Boundary Detection Techniques," *IS&T/SPIE*, vol. 2670, pp. 170-179, Feb. 1996.
- [10] B. Shahraray, "Scene Change Detection and Content-based Sampling of Video Sequences," *Proc. SPIE*, vol. 2419, pp. 2-13, Feb. 1995.
- [11] 이재연, 전병태, 배영래, "복수의 검출기를 사용하는 동영상의 샷 경계 자동 검출 알고리즘," vol. 5, no. 9, pp. 1352-1360, *정보과학회 논문지*, 1998.
- [12] Simon Haykin, Neural network, Macmillan College Publishing Compay, 1994.
- [13] Joan L. Mitchell, William B. Pennebaker, Chad E. Fogg, and Didier J. LeGall, MPEG Video Compression Standard, Chapman & Hall, 1996.

저 자 소 개

전 병 태



- 1986년 2월 : 한남대학교 전산과(학사).
- 1989년 8월 : 숭실대학교 전산과(석사).
- 1997년 3월 ~ 현재 : 고려대학교 컴퓨터학과 박사과정(박사수료).
- 1992년 5월 : IR52 장영실 상 수상(과기부 장관상).
- 1989년 9월 ~ 현재 : 한국전자통신연구원 컴퓨터·소프트웨어 기술연구소 선임연구원 재직중.
- 주관심분야 : 동영상 처리, 컴퓨터 비전, 멀티미디어 영상처리 등



이 재 연

1984년 : 서울대학교 이학사(제어계측)

1986년 : 한국과학기술원 공학석사(전기및전자)

1992년 : 일본 東海대학 공학박사(광공학)

1986년 3월 ~ 현재 : 한국전자통신연구원 컴퓨터·소프트웨어 기술연구소 책임연구원 재직중

주관심분야 : 동영상 처리, 멀티미디어 영상처리 등



배 영 래

1976년 : 서울대학교 해양물리전공(이학사).

1986년 : 한양대학교 전자계산 전공(공학석사).

1992년 : Univ. of Kent 전자공학 전공(공학석사).

1995년 : Univ. of Kent 전자공학 전공(공학박사).

1976년 ~ 1979년 : 해군장교.

1979년 ~ 1980년 : 삼성.

1980년 : KIST입소, 현재 한국전자통신연구원 컴·소 연구소 영상정보처리연구팀장.

주관심분야 : 영상처리, 컴퓨터 비전, 병렬처리