논문-03-08-3-10

주파수 분석을 통한 인버스 텔레시네 기법

구 형 일*, 조 남 익*, 이종원**

Inverse Telecine by Using Frequency Analysis

Hyung-Il Koo*, Nam-Ik Cho* and Jong-Won Lee**

요 약

초당 24프레임으로 제작된 영화나 광고물 등을 TV로 방영하거나 비디오테이프로 제작할 때에는 몇 개의 필드를 반복하여 사용함으로써 초당 60 필드로 만드는데 이를 텔레시네 또는 3:2 풀다운이라 한다. 비디오테이프나 TV 방송을 MPEG으로 인코딩할 때 텔레시네 영상물에 대해서는 이를 다시 초당 24 프레임의 순차주사 영상으로 바꾸어 인코딩하면 일반적으로 60 필드를 모두 인코딩하는 경우에 비하여 20% 정도의 비트를 절약할 수 있다. 본 논문에서는 텔레시네 영상의 특성을 이용하여 인버스 텔레시네를 수행하는 알고리듬을 제안하였다. 구체적으로, 본 방법은 텔레시네 과정에서 다른 시간에 촬영된 필드가 합쳐지면 홀수줄과 짝수줄의 불일치가 생기고 이는 나이퀴스트 주파수 부근에서 큰 성분을 갖는다는 것을 이용한다. 실험결과, 기존의 움직임을 이용하는 방법은 계산량이 많고 VHS와 같이 화질이 높지 않은 경우에는 잘 동작하지 않는 반면에 제시된 방법은 화질에 관계없이 더 높은 신뢰도와 간단한 계산으로 인버스 텔레시네를 수행함을 확인하였다.

Abstract

When a cinema being composed of 24 frames/sec is converted to NTSC video or TV program, several fields are repeated to have 60 fields/sec, which is called the telecine or 3:2 pull-down. Hence, when encoding the telecine NTSC video into MPEG format, if we convert it into the original 24 frames/sec progressive cinema, then we can save 20% of bits compared to the case of encoding all the 60 fields. In this paper, we propose an algorithm for performing the inverse telecine by using the properties of the frames. Specifically, the algorithm exploits the fact that there is much inconsistency between the even and odd fields in the case of telecine frame, which results in high magnitude at the Nyquist frequency in the vertical direction. The experiment shows that the proposed algorithm performs very well regardless of the quality of video with a very few computations, whereas the conventional motion based method requires much computational complexity and its performance is degraded when the video is of low (eg. VHS) quality.

I. 서 론

영화나 광고물은 초당 24 프레임으로 촬영되고 저장되므로, 이를 TV로 방영하거나 비디오테이프에 저장하기 위해서는 초당 30 프레임, 즉 초당 60 필드로 바꾸어주어야 한다. 이렇게 초당 24 프레임을 초당 60 필드로 바꾸어 주는 작업을 텔레시네 혹은 3:2 풀다운(pull-down)이라고 한다[1][2][3]. 이 작업은 2 개의 필름 프레임에서 1 개의 필드를 반복해서 5 개의 필드를 만들어주는 작업으로 요약할 수 있다. 그림 1은 텔레시네 과정을 보여 주고 있다. 여기서 4

* 서울대학교 전기공학부, 뉴미디어통신공동연구소
School of Electrical Engineering Seoul National University
** 삼성전자, DS총괄, System LSI사업부
Samsung Electronics, Device Solution Network, System LSI

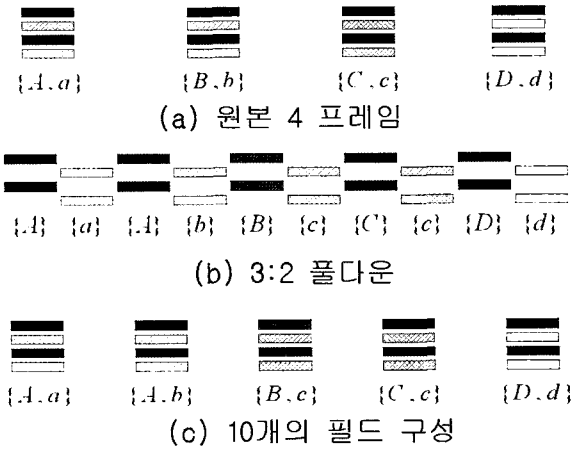

그림 1. 텔레시네 과정 (탑필드 우선 방식)
Fig 1. Telecine procedure (top field first)

장의 원본 프레임을 그림 (a) 에서와 같이 $\{A,a\}$, $\{B,b\}$, $\{C,c\}$, $\{D,d\}$ 라고 표기하고 있으며 대문자는 탑필드(top field)를 나타내는 기호이고, 소문자는 바텀 필드(bottom field)를 나타내는 기호이다. 그림 (b)는 첫째 셋째 프레임으로부터 두 필드를 더 만들어내는 과정이며 이를 둘씩 짝지어 결국 그림 (c)와 같이 $\{A,a\}$, $\{A,b\}$, $\{B,c\}$, $\{C,c\}$, $\{D,d\}$ 인 5 장의 프레임이 나오게 되고 이를 텔레시네 영상물이라고 한다. 3:2 라는 말은 그림 1 에서 볼 수 있듯이 첫 번째와 세 번째 프레임에서는 각각 3 개의 필드를 만들어주고, 두 번째와 네 번째 프레임에서 각각 2 개의 필드를 만들기 때에 붙여진 이름이다. 이상은 탑 필드 우선 방식(top field first)에 대한 설명이며, 바텀 필드를 먼저 복사하는 바텀 필드 우선 방식도 존재한다.

TV나 비디오를 디지털 미디어에 저장할 때, 신호 그대로 초당 60 필드 인터레이스 방식으로 저장할 수도 있지만 텔레시네 영상물의 경우 초당 60 필드 인터레이스를 역으로 초당 24 프레임 프로그레시브로 바꾸어 저장하면 비트율이 약 80 % 로 줄어든다는 장점이 있다. 여러 가지 영상에 대한 실험을 해 보았을 때, 약간의 차이를 보이기는 하여도 대부분의 경우는 그림 2와 같이 모든 비트율에 대하여 실제로 비트양이 20% 정도 줄어든다. 이와 같이 인버스 텔레시네를 한 후에 저장된 MPEG 비디오 등은 헤더에 24 프레임 순차주사로 저장된 것으로 표시하여 플레이어는 출력할 때, 다시 3:2 풀다운을 수행하여 TV에서 볼 수 있도록 해 줄 수 있다. 이와 같은 이점을 얻기 위해서 초당 60

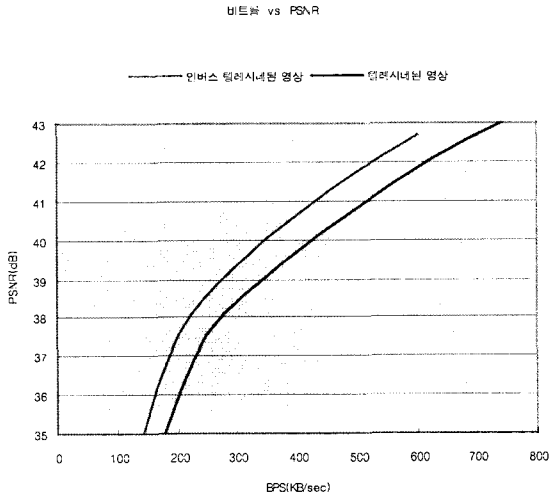

그림 2. 텔레시네 영상과 인버스 텔레시네된 영상의 비트율과 PSNR 그래프
Fig 2. Comparison of PSNR of telecined sequence and inverse telecined sequence

필드 인터레이스 방식으로 만들어진 텔레시네 영상물을 다시 원래의 초당 24 프레임 프로그레시브로 바꾸는 작업을 인버스 텔레시네라 한다. 편의상, 앞에서 언급된 영상과 앞으로 설명될 영상의 종류를 정리하면 다음과 같다.

- **텔레시네 영상** : 원래 초당 24 프레임으로 제작되었고 텔레비전 방송을 위해 필드를 반복하는 방법으로 초당 60 필드로 바꾼 영상
- **텔레비전 영상** : 방송용으로 처음부터 초당 60 필드로 제작된 영상
- **혼합 영상** : 텔레시네 영상과 텔레비전 영상이 혼합된 경우 (텔레시네 영상에 자막 등의 그래픽을 넣은 경우)

앞에서 언급된 바와 같이 텔레시네 영상의 경우는 인버스 텔레시네를 수행함으로써 압축률을 높일 수 있다. 그러나 텔레비전 영상이나 혼합 영상의 경우 인버스 텔레시네를 수행하면 필드의 손실을 초래하게 되므로 인버스 텔레시네를 수행하면 안 된다. 따라서 인버스 텔레시네를 수행하기 위해서는 텔레시네 영상과 나머지 둘을 구별해내는 기술이 필요하다. 여기서 문제가 되는 것은 대부분의 혼합 영상의 경우 텔레비전 영상 요소가 화면에서 적은 면적만을 차지하여 화면의 전체적인 특성이 텔레시네 영상의 형태를 보인다는 점이다. 실제로 기존 방법[4]은 대부분의 혼합 영상을 텔레시네 영상으로 인식하고 있다. 또한, 텔레시네가 항상 규칙성 있게 수행되는 것은 아니고 장면전환이

나 프레임 유실로 그 규칙성이 깨질 수 있으므로 이러한 예외적인 경우를 포함하여 3:2 풀다운이 이루어진 방법을 추정함으로써 원본 영상을 복원해 주어야 한다.

이상의 작업 들은 원칙적으로 필드 사이의 SAD(sum of absolute difference)를 구하는 등의 간단한 방법으로 해결 가능하지만 실제로는 잡음의 영향이 있으므로 좀 더 정교하고 세밀한 방법이 필요하다. 즉 잡음이 추가된 경우 두 필드가 공간적 시간적 인접성 때문에 작은 SAD값을 보이는 것인지 아니면 같은 필드에 잡음이 추가되었기 때문에 작은 SAD 값을 보이는 것인지 판단하기 어렵다. 또한 텔레시네의 디스플레이 규칙은 잃어버린 프레임이나 장면 전환에 따라 변할 수 있다. 그리고 이런 디스플레이 규칙은 경우에 따라 매우 빈번하게(수십 프레임에 한 번씩) 변할 수 있으므로, 디스플레이 규칙의 변화를 빨리 찾고 즉각적으로 반응하는 알고리듬이 필요하다.

이러한 문제를 해결하기 위한 기존의 인버스 텔레시네 기법으로는 움직임 벡터를 이용한 방법이 있다[1]. 이 방법은 반복된 필드 사이에서는 잡음의 영향이 있더라도 움직임 벡터의 크기가 매우 작게 나온 다는 점을 이용했다. 구체적으로 그림 1-(c)의 $\{A,a\}$, $\{A,b\}$, $\{B,c\}$, $\{C,c\}$, $\{D,d\}$에서 첫 번째 탑 필드와 두 번째 탑 필드 사이의 움직임 벡터의 크기는 첫 번째 탑 필드와 세 번째 탑 필드 사이의 움직임 벡터의 크기보다 매우 작을 것이다. 마찬가지로 첫 번째 바텀 필드와 두 번째 바텀 필드 사이의 움직임 벡터에 비해서도 매우 작은 값을 기대할 수 있을 것이다. 따라서 이 방법은 움직임 벡터의 크기의 비가 어떤 문턱치를 넘는가 아닌가를 판별하여 텔레시네를 검출하게 된다.

이 방법의 문제점은 우선 문턱치가 두 개 필요하다는 것이다. 탑 필드와 바텀 필드의 문턱치는 잡음의 양이 많은 영상(극단적인 경우에는 VHS VTR)인가 아닌가에 따라 크게 변한다. 화질에 따른 자동적인 문턱치 선정은 매우 어려우므로 다양한 화질의 영상에 대해 골고루 좋은 성능을 내기에는 문제가 있다. 그 다음 문제점은 움직임 벡터를 구하기 위해서는 많은 계산량이 필요하다는 것이다. 경우에 따라 구한 움직임 벡터를 재활용할 수도 있지만, 때로는 필요가 없는 움직임 벡터를 구해야하거나 필드가 합쳐져 새로운 프레임이 나옴에 따라 움직임 벡터를 새로 구해야하는 일이 발생하게 된다. 또한 이 방법은 전체적인 구조를 고려하지 않고 있다. 텔레시네는 비록 장면 전환이나 다른 원인으로 인해 디스플레이 규칙이 변하지만 전체적으로 보아 다섯 프레임 단위로 디스플레이 규칙이 강한 연관성을 보인다. 즉 앞의 다섯 프레임이 $\{A,a\}$, $\{A,b\}$, $\{B,c\}$, $\{C,c\}$, $\{D,d\}$ 의 구조를 보이고 있다면 그 후의 다섯 프레임도 위의 구조를 보일 확률이 매우 높다. 하지만 [1]의 방법에서는 전후 관계에 대한 고려가 전혀 없다. 따라서 전후 그룹의 관계를 고려함으로써 더 좋은 성능을 낼 수 있는 여지가 남아 있다. 마지막으로 이 방법은 화면의 전체적인 특성만을 고려하므로 혼합 영상을 텔레비전 영상이 아니라 텔레시네 영상으로 파악하게 된다는 문제점이 있다.

위에서 언급된 문제점들을 개선하고 인버스 텔레시네의 정확도를 높이기 위해서 본 논문에서는 특정 주파수 성분값과 화면의 일부 만을 이용하는 변형된 SAD(sum of absolute difference)를 이용하여 영상의 종류를 판별하고 인버스 텔레시네를 수행하는 방법을 제안한다. 또한 이 방법은 이전 프레임과의 관계를 고려함으로써 거의 모든 화질에서 높은 정확도를 보인다.

본 논문의 구성은 다음과 같다. 제 1 장 서론에 이어, 제 2 장에서는 제안된 알고리듬이 소개되고, 제 3 장에서는 여러 종류의 영상에 대한 실험 결과가 소개된다. 마지막으로 제 4장에서 결론을 맺는다.

Ⅱ. 제안된 방법

앞에서 언급된 바와 같이 텔레시네 영상은 두 종류의 프레임으로 나눌 수 있다. 즉 $\{A,b\}$, $\{B,c\}$처럼 시간 간격(1/24 초 차이)을 두고 촬영된 탑 필드와 버텀 필드가 합쳐져 만들어진 프레임이 있고, 같은 시간에 촬영된 필드로 이루어진 $\{A,a\}$, $\{C,c\}$, $\{D,d\}$ 같은 프레임으로 나눌 수 있다. 정리하면, 프레임은 다음과 같은 두 가지 형태로 구분할 수 있다.

0번 형태의 프레임:$\{A,a\},\{C,c\},\{D,d\}$
1번 형태의 프레임:$\{A,b\},\{B,c\}$

다른 시간대에 촬영된 두 필드가 합쳐져서 프레임을 구성하면(1번 형태) 짝수줄과 홀수줄이 그림 3-(a)에서와 같이 서로 엇갈려 나타난다. 그림 4-(a)와 같이 같은 시간대의 필드를 합하여 구성된 프레임(0번 형태)의 경우, 1번 형태의 경우보다 짝수줄과 홀수줄의 엇갈림이 작다는 것을 볼 수 있다. 이러한 엇갈림은 세로 방향으로 보았을 때, 매 픽셀

(a) 1 번 형태의 프레임 (b) 푸리의 변환의 크기

그림 3. 1 번 형태의 프레임과 이의 푸리에 변환의 크기

Fig 3. Type 1 frame and the magnitude of its Fourier transform

(a) 0 번 형태의 프레임 (b) 푸리의 변환의 크기

그림 4. 0 번 형태의 프레임과 이의 푸리에 변환의 크기

Fig 4. Type 0 frame and the magnitude of its Fourier transform

단위로 변화가 있는 것이므로 주파수 영역에서 가장 높은 주파수인 π 부근의 주파수 성분을 크게 한다. 그 예로 그림 3, 4의 (b)에는 각 영상의 푸리에 변환의 크기를 보였다. 중앙의 흰색 부분이 저주파수의 세기이며, 중앙의 맨 위와 아래가 세로 방향으로 π 부근의 주파수 세기이다. 그림 4의 0번 형태의 경우 미미한 세기로 있던 고주파 영역이 그림 3의 1 번 형태의 프레임에서는 매우 큰 값을 보임을 알 수 있다.

그러나 영상에 따라 원래 π 성분의 주파수 성분이 강한 영상일 수도 있으며 그런 경우에는 고주파 성분이 줄어 들 수도 있다. 하지만 이 경우에도 1번 형태의 프레임은 0 번 형태의 프레임과 다른 특성의 고주파 성분을 갖는다. 즉 다섯 장이 연달아 있다면 세 장은 비슷한 수준의 고주파 세기를 보일 것이며, 두 장은 거기에서 이탈된 주파수 값을 보일 것이다. 따라서 고주파 성분의 크기를 특징값으로 했을 때 0 번 형태 프레임의 특징값을 0에 가깝도록 보내고, 1번 형태 프레임의 특징값을 1 가까이 보내는 정규화 과정을 통해, 프레임들을 0과 1 사이의 실수의 수열로 보면서 1 번 형태의 프레임을 찾는 동시에 프레임을 잃어버리거나 디스플레이 규칙이 변하는 것을 관찰할 수 있다.

이렇게 주파수 성분을 관찰함으로써 같은 것으로 추정되는 두 개의 필드의 위치를 파악할 수 있다. 만약 입력 영상이 텔레시네된 것이라면 이 결과는 실제로 반복된 필드를 추정해 낼 것이고 텔레비전 영상인 경우 의미 없는 위치를 찾아 낼 것이다. 이렇게 추정된 위치에서 필드의 일치-불일치를 판정함으로써 실제로 영상이 텔레시네 영상인지 텔레비전 영상인지 결정하게 된다.

요약하자면, 각 프레임의 주파수 성분의 상호 관계를 관찰해서 같다고 추정되는 필드의 위치를 파악하고 실제로 같다면 텔레시네 영상, 같지 않다면 텔레비전 영상이라고 판단하는 것이다. 여기서 필드의 일치와 불일치는 화면의 8 픽셀을 하나로 묶어서 생각하며 화면의 오직 일부분만을 고려하기 때문에 혼합 영상의 경우에도 문제가 되지 않는다.

제안된 알고리듬의 수행 과정은 다음과 같이 템플릿 비교, 영상물 종류 판단, 복원 세 단계로 나누어진다.

1. 템플릿 비교기

1.1 고주파 성분 추출

세로 방향의 고주파 성분의 값을 얻는 방법은 화면을 커다란 블록 (예를 들어, 128×128)으로 나누고, 각 블록을 푸리에 변환을 한 후, 세로 방향으로 π 부근의 주파수 성분의 값을 더하는 것이다. 그러나 이 방법은 계산량이 많다는 문제점이 있다. 따라서 본 논문에서는 커다란 블록으로 나누는 방법보다는 주파수 성분의 값을 정확하게 계산하지 못하더라도 계산량이 적은 방법을 사용한다. 적은 양의 계산으로 π 성분 값을 추정하는 방법은 화면을 작은 블록 (4×4)으로 나눈 다음 각 블록에서 π 성분 값을 구하고

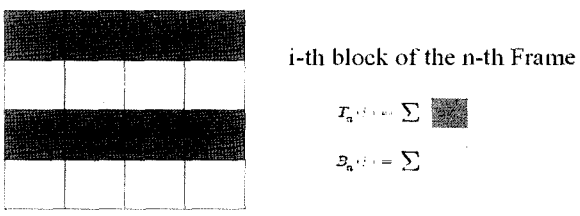

그림 5. 화면의 단위 블록의 특징 값

Fig. 5. The characteristic value of each block

표 1. 디스플레이 규칙을 찾기 위하여 사용된 템플릿
Table 1. Templates used for finding the display rule

번호	템플릿	새로 받아들일 프레임 수
1	{1,1,0,0,0}	5
2	{0,1,1,0,0}	1
3	{0,0,1,1,0}	2
4	{0,0,0,1,1}	3
5	{1,0,0,0,1}	4
6	{1,1,0,0,1}	4
7	{1,0,0,1,1}	3
8	{1,1,0,1,1}	3
9	{1,0,1,1,0}	2
10	{1,0,0,0,0}	5
11	{0,0,0,0,0}	5

이 값들을 모든 블록에 걸쳐 더하는 것이다. n번째 프레임의 i번째 블록이 그림 5처럼 생겼다면 위에서 설명한 값은 아래와 같다.

$$H_n = \sum_i | T_n(i) - B_n(i) | \tag{1}$$

길이 N인 디지털 신호 x_n의 푸리에 변환에서 π성분의 값은

$$X(\pi) = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N} \cdot n \cdot \frac{N}{2}} \tag{2}$$

이므로[5], 결국 이는 $\sum_{n=0}^{N-1} x_n (-1)^n$ 으로서 위의 식 (1)에서와 같이 한 픽셀씩 더하고 빼는 것과 같다. 따라서 모든 블록에 대한 위 식의 합은 푸리에 변환에서 π 부근의 주파수 성분 값을 나타내며 그 프레임의 주파수 특징 값이 된다.

1.2 정규화

위에서 추출한 특징값은 화면의 영상 수준, 동작의 양 그리고 원본 영상의 π 주파수 수준에 따라 각기 다른 크기의 값을 보이므로 정규화 과정이 필요하다. 만약 5 프레임에 걸쳐 특징값을 추출한다면 5 개중 3 개는 중간 수준의 값(0번 형태)을 보이고, 나머지 2 개(1번 형태)는 중간에서 이탈한 값을 보일 것이다. 정규화의 목적은 중간 수준의 값을 0 으로 보내고 이탈된 값을 1로 보내는 것이다. 따라서 정규화 식은 기본적으로

$$\frac{x_i - median}{2 \times \sigma} \tag{3}$$

의 꼴이다. 2σ 의 값으로 나누는 이유는 가우시안 이라고 생각하였을 때 95% 정도를 0과 1사이에 넣기 위해서이다. 여기서 x_i는 각 프레임의 특징값, $median$은 다섯 개 값의 중간 값, σ 는 표준편차이다. 그러나 이 결과가 0과 1 사이의 값에 모두 들어오도록 클리핑을 해주어야 하고 표준편차가 너무 작으면 의미 없는 값들이 큰 폭으로 변할 수 있으므로, 분모가 얼마 이상의 값만 들어올 수 있도록 하기 위해서 정규화 식은 결국 다음과 같이 정의해야 한다.

$$clipping\left(\frac{x_i - median}{2 \times \max(\sigma, \sigma_0)} \right) \tag{4}$$

여기서 σ_0 는 σ가 작을 때 대체되는 상수이다.

1.3 템플릿 비교

정규화를 통해 나온 다섯 개의 수열을 관찰하여 프레임들이 어떤 상태로 있는지 판단한다. 1번 형태의 프레임은 정규화 값이 1 에 가까운 값을 보일 것이고, 0 번 형태의 프레임은 0 에 가까운 값을 보일 것이다. 따라서 이상적인 경우라면 1번 형태 프레임의 시작 위치에 따라 {1,1,0,0,0}, {0,1,1,0,0}, {0,0,1,1,0}, {0,0,0,1,1}, {1,0,0,0,1} 의 다섯 가지 템플릿과 비교하면 된다. 하지만 장면 전환이나 잃어버린 프레임으로 인해, {1,1,0,0,1} 등의 상황이 발생할 수 있으므로 모든 경우를 포함하기 위한 템플릿의 수는 1과 0을 5자리에 늘어놓는 경우의 수($2^5 = 32$)이다. 그러나 모든 가능한 경우와 비교하는 것은 정확한 판단을 내리는데 방해가 될 뿐만 아니라 시스템 자원의 낭비이다. 만약 지난 다섯 프레임의 판단 정보로부터 현재 템플릿이 {1,1,0,0,0} 이기대 되도록 만든다면 비교해주어야 할 대상이 훨씬 줄어들며 판단도 내리기 쉬울 것이다. 예를 들어, 현재 수열이 {1,1,0,0,1} 이라고 판단이 내려진다면, 디스플레이 규칙이 변했거나, 0 번 형태의 프레임 한 장이 빠졌다고 생각하고, 앞의 넉 장을 내보내고 네 장만 더 받아들여서 다음 상태가

{1,1,0,0,0} 이 기대되도록 만드는 것이다. 표 1은 사용된 템플릿과 각 경우에 대해 다음 차례에 {1,1,0,0,0}이 기대되도록 새로 받아들여야할 프레임의 수이다. 표 1에서 (10), (11) 항목은 예외적인 템플릿으로서 1번 템플릿의 변형으로 움직이는 물체가 화면 밖으로 빠져나가거나 동작이 거의 없는 경우에 나타나는 수열이다.

1.4 판단

지금 막 인코딩이 시작되어 과거 프레임에 대한 정보가 전혀 없다면 위에서 내린 판단을 그대로 따르면 된다. 하지만 과거 프레임에 대한 정보가 있다면 좀 더 정교한 판단을 거쳐야 한다. 즉 위에서 언급한 템플릿들과 단순한 거리 비교를 하여 만약 표 1의 (1), (10), (11) 항목과 가장 가깝다고 결과가 나왔다면 이전 결정과 일관성이 있는 결정이므로, 그 추측이 옳다고 판단할 수 있다. 하지만 이외의 템플릿과 가장 가깝다면, 이 결과가 잡음이나 다른 영향에 따른 잘못된 결과인지, 아니면 장면 전환이나 텔레시네 알고리듬 자체의 잘못으로 발생한 변화에 의한 결과인지 판단을 내려야한다.

만약 가장 가까운 템플릿이 표 1의 (1), (10), (11) 항목이 아니라면 판단의 수용 여부는 두 가지 기준에 의해 결정된다. 첫 번째 기준은 모든 거리들이 비슷한 크기를 보이며 모여 있는가, 아니면 뚜렷이 다른 값을 가지며 흩어져 있는 가이다. 일반적으로 판단을 내리기 어려운 경우는 정규화 된 값들이 0.5 전후를 보여서 0번 형태의 프레임인지 1번 형태의 프레임인지 판단하기 어려운 상황이다. 이런 상황에서는 작은 차이로 새로운 결정을 내리는 것보다 기존 결정을 그대로 유지하는 것이 유리하다. 정리하면, 새로운 결정을 받아들이기 위한 조건 중 하나는 아래와 같이 주어진 수열과 템플릿의 여러 수열과의 거리의 표준 편차가 어떤 문턱치 이상이 되는 경우로 해야 한다.

$$\sigma_{\text{템플릿과의 거리}} > \tau_1 \tag{5}$$

두 번째 기준은 (1)번 템플릿과의 거리와 가장 짧은 거리가 얼마나 차이가 나는가 하는 것이다. 기존 결정을 유지할 지, 새로운 결정을 내려야하는지를 결정해야 하는 상황에서 필요한 정보이다. 즉, 새로운 결정을 받아들이기 위한 또 한 가지 조건은

$$(1)\text{번 템플릿과의 거리} - \text{최소 거리} > \tau_2 \tag{6}$$

이다. 이상의 두 가지 조건을 만족하면 새로운 결정이 받아들여진다. 그렇지 않으면 기존 결정이 유지된다. 이상의 작업으로 다섯 개의 프레임이 들어왔을 때, 이 프레임을 표 1에 나타난 다섯 개의 템플릿 중에 하나로 파악하게 된다.

2. 영상물 종류(텔레시네/텔레비전/혼합)의 판단

템플릿 수열의 종류 중에서 텔레시네 영상에서 빈번하게 나오는 수열은 {1,1,0,0,0}, {1,0,0,0,0}, {0,0,0,0,0} 이다. 이들을 정상 텔레시네 수열이라고 하고 나머지를 비정상 텔레시네 수열이라고 하자.

2.1 정상 텔레시네 수열 비율 검사

만약 텔레시네 영상이라면 대부분의 수열은 정상 텔레시네 수열일 것이다. 반면 정상 텔레시네 수열의 비율이 작다면 (예를 들어 절반 이하) 이는 텔레비전 영상이나 혼합 영상일 것이다. 따라서 정상 텔레시네 수열의 비율로부터 일차적으로 인버스 텔레시네 해주지 말아야 하는 영상을 분류해낸다.

2.2 특징 추출

그림 5의 기호를 그대로 사용하여 다음과 같은 특징값을 구한다. 여기서 $k-largest$의 의미는 Σ 내부의 값들 중에서 큰 k개의 값을 더한 다는 뜻이다. 이 값은 의미상 여덟 개의 픽셀을 하나로 묶은 SAD(sum of absolute difference)이다. 하지만 여덟 개의 픽셀을 하나로 묶었다는 점에서 잡음의 영향을 현저히 줄인 값이며 화면 전체를 고려하지 않고 차이가 큰 일부분만을 고려했다는 점에서 화면의 일부분에서 발생하는 특징도 제대로 검출해내게 된다.

$$TM_n = \sum_{k-largest} | T_n(i) - T_{n-1}(i) | \tag{7}$$

$$BM_n = \sum_{k-largest} | B_n(i) - B_{n-1}(i) | \tag{8}$$

TM_n은 n번째 프레임과 $n-1$번째 프레임의 탑 필드의 불일치를 나타내는 지수이며, BM_n 은 n번째 프레임과 $n-1$번째 프레임의 바텀 필드 사이의 불일치를 나타내는 값이다.

2.3 영상물 종류 판단

현재의 다섯 프레임들에 대한 템플릿 비교기의 결과가 {1,1,0,0,0}, {1,0,0,0,0}, {0,0,0,0,0} 의 수열 중에 하나였다면 탑 필드 우선 방식 텔레시네, 바텀 필드 우선 방식 텔레시네, 텔레비전 영상인가에 따라서 표 2의 세 가지의 필드 구조 중에 하나를 가지게 된다. 표 2의 가로 방향은 프레임의 번호를 나타낸다. 현재의 다섯 프레임과 이전 한 프레임을 포함해서 총 여섯 개의 프레임의 구조를 나타내고 있다. 세로 방향은 탑 필드와 바텀 필드를 의미한다. 같은 필드는 같은 문자로 표시되어 있다. 따라서 다음의 값들은 영상의 종류에 따라 각기 다른 성질의 보이게 된다.

$$\log\frac{TM_1}{BM_1} \tag{9}$$

$$\log\frac{BM_3}{TM_3} \tag{10}$$

구체적으로 탑 필드 우선 방식(표 2-a)이었다면 TM_1은 A와 A의 불일치이며, BM_1은 a와 b의 불일치이다. 마찬가지로 TM_3은 B와 C의 불일치이며 BM_3은 c와 c의 불일치이다. 따라서 식 (9)와 식 (10) 모두 큰 음수 값을 보이게 된다. 마찬가지 논리로부터 바텀 필드 우선 방식은 모두 큰 양수의 값을 보인다. 그리고 텔레비전 영상이나 혼합 영상의 경우 0에 가까운 값을 보인다. 따라서 이 값을 가지고 영상물의 종류를 판별할 수 있다.

표 2. 영상의 종류에 따른 프레임의 구조
Table 2. Structure of frames according to their types

(a) 탑필드 우선 방식의 텔레시네 영상

	0	1	2	3	4	5
탑	A	A	B	C	D	E
바텀	a	b	c	c	d	e

(b) 바텀 필드 우선 방식의 텔레시네 영상

	0	1	2	3	4	5
탑	A	B	C	C	D	E
바텀	a	a	b	c	d	e

(c) 텔레비전 영상

	0	1	2	3	4	5
탑	A	B	C	D	E	F
바텀	a	b	c	d	e	f

2.4 판단 기준

템플릿 비교기를 통해 같다고 생각되는 필드의 위치를 파악하였지만 템플릿 비교기의 결과가 잘못되었을 수도 있고 프레임의 소실로 인해 위의 가정이 무너질 수도 있다. 따라서 식 (9)와 (10)의 값을 몇 개 누적해서 사용한다면 약간의 지연을 불러일으킬 수 있지만 신뢰할 만한 값을 얻을 수 있다. 그리고 텔레시네 영상을 인버스 텔레시네하지 않는 것은 정보 측면에서 손실이 없지만 텔레비전이나 혼합 영상을 인버스 텔레시네하는 것은 정보의 손실을 가져오므로 동작이 강하고(인버스 텔레시네한다면 손실되는 정보가 큰 경우) 텔레비전 영상일 가능성이 높다면 즉각적으로 영상을 인버스 텔레시네하지 않는 것이 옳다. 따라서

표 3. 템플릿에 따른 복원 방법
Table 3. Reconstruction methods for each template.

번호	템플릿	방출하는 프레임 수	복원 방법
1	{1,1,0,0,0}	4	1, 2 번째 프레임을 합치고, 3, 4, 5 번째 프레임 방출
2	{0,1,1,0,0}	1	1 번째 프레임 방출
3	{0,0,1,1,0}	2	1, 2 번째 프레임 방출
4	{0,0,0,1,1}	3	1, 2, 3 번째 프레임 방출
5	{1,0,0,0,1}	3	1 프레임 삭제 후, 2, 3, 4 번째 프레임 방출
6	{1,1,0,0,1}	4	1, 2 번째 프레임을 합치고, 3, 4 번째 프레임 방출
7	{1,0,0,1,1}	2	1 번째 프레임 삭제 후, 2, 3 번째 프레임 방출
8	{1,1,0,1,1}	2	1, 2 번째 프레임을 합치고, 3 번째 프레임 방출
9	{1,0,1,1,0}	1	1 번째 프레임 삭제 후, 2 번째 프레임 방출
10	{1,0,0,0,0}	4	1 번째 프레임 삭제 후 2, 3, 4, 5 번째 프레임 방출
11	{0,0,0,0,0}	4	1, 2 번째 프레임을 합치고, 3, 4, 5 번째 프레임 방출

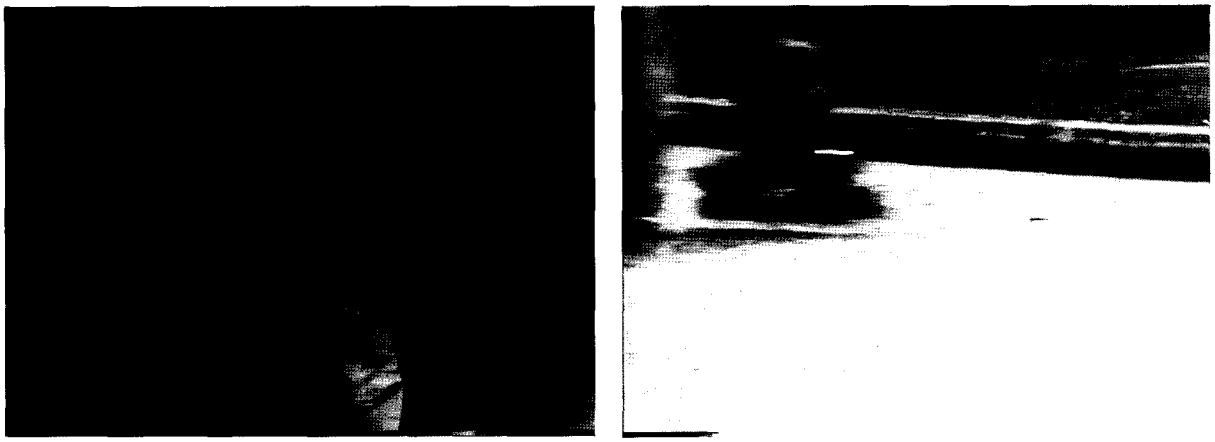

그림 6. 테스트 영상 1 (친니친니)
Fig 6. Test Sequence 1 (Kissing you)

그림 7. 테스트 영상 2 (제5원소)
Fig 7. Test Sequence 2 (The fifth element)

그림 8. 테스트 영상 3 (매트릭스)
Fig 8. Test sequence 3 (Matrix)

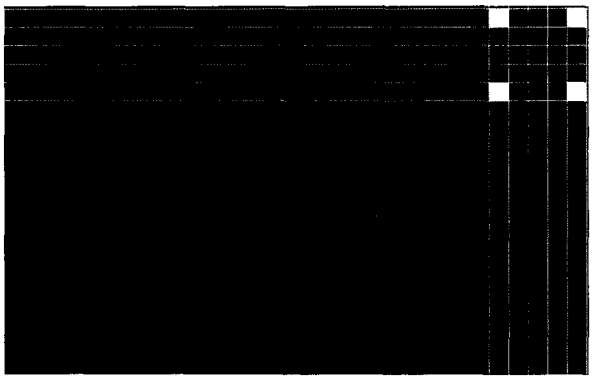

그림 9. 테스트 영상 4
Fig 9. Test sequence 4

그림 10. 테스트 영상 5
Fig 10. Test sequence 5

그림 11 테스트 영상 12
Fig 11. Test sequence 12

TM_1, BM_1, TM_3, BM_3 의 값이 모두 크고 식 (9)와 (10)의 값이 0 에 가깝다면 즉각적으로 인버스 텔레시네 중지하라는 신호를 내보내도록 한다.

3. 복원

템플릿 비교기의 결과와 영상 종류 판별기의 결과를 통해 영상을 재구성할 수 있다. 텔레비전 영상(또는 혼합 영상)으로 판단된 경우 그대로 초당 60 필드로 저장하고, 텔레시네 영상으로 판단된 경우 반복된 필드를 삭제하면 된다. 비정상 텔레시네 수열인 경우 복원 방법은 여러 가지가 있을 수 있는데 하나의 예가 표 3 에 나타나 있다. 1 프레임과 2 프레임을 합친다는 의미는 탑 필드 우선 방식의 경우 2 번 프레임의 탑 필드와 1 번 프레임의 바텀 필드를 합쳐서 하나의 프레임의 만들며, 바텀 필드 우선 방식의 경우 1 번 프레임의 탑 필드와 2 번 프레임의 바텀 필드를 합쳐서 하나의 프레임을 만드는 것을 의미한다.

Ⅲ. 실험 결과

실험에 사용된 영상과 그 특징은 표 4에 나와 있다. 실험은 영상 종류 판단과 텔레시네 영상의 역변환 두 가지로 나누어 행하였다.

표 4. 실험에 사용된 영상과 그 특성
Table 4. Moving pictures used in the experiment and their characteristics.

	영상 종류	이름	설명
1	TF 텔레시네	친니친니	VTR에서 캡쳐된 저화질 영상
2	TF 텔레시네	제5원소	VTR에서 캡쳐된 저화질 영상
3	TF 텔레시네	매트릭스	합성된 텔레시네 영상
4	BF 텔레시네	폐루자 영상	비교적 고화질인 텔레시네 영상
5	BF 텔레시네	보드 영상	비교적 고화질인 텔레시네 영상
6	텔레비전 영상	BBC	
7	텔레비전 영상	Flower Garden	
8	텔레비전 영상	Susie	
9	혼합 영상	제5원소(Co)	제5원소 중간 100 프레임에 컴퓨터 그래픽 추가
10	혼합 영상	제5원소(Ca)	제5원소 중간 100 프레임에 흘러가는 자막 추가
11	혼합 영상	보드영상(Co)	보드영상 중간 100 프레임에 컴퓨터 그래픽 추가
12	혼합 영상	보드영상(Ca)	보드영상 중간 100 프레임에 흘러가는 자막 추가

1. 영상 종류 판단 실험

다른 종류의 실험 영상을 100 프레임 단위로 이어서 만든 1200 프레임의 영상에 대해서 필름 모드를 제대로 찾아내는가를 실험하였다. 그리고 표 4의 (9), (10), (11), (12) 항목의 영상에 대해서 컴퓨터 그래픽이나 흘러가는 자막이 나오는 순간과 사라지는 순간에 영상 종류에 대한 판단이 변하는지 관찰하였다. 식 (7), (8)에서 k는 3으로 하였다.

실험 결과 9 번의 영상 종류 변화 중에서 4 회는 즉각적인 반응을 보였고, 5 회는 5에서 10 프레임씩 지연이 있었다. 문제가 되는 것은 텔레비전 영상을 텔레시네 영상으로 인식하여 반복되지 않은 필드를 삭제하는 것인데, 이 관점에서 1200 프레임 중에서 잘못된 삭제는 3 필드만 일어났다. 표 4의 (9), (10), (11), (12) 항목의 영상의 경우 텔레시네 모드에서 혼합 모드로 변하는 순간 영상 모드가 즉각적으로 변했으며, 혼합 모드에서 텔레시네 모드로 돌아오는데 각각 5 프레임의 지연이 있었다.

2. 텔레시네 역변환

실험은 [1] 방법과 제안된 방법에 대해서 수행하였다. (5), (6) 식에서 언급된 상수는 $\sigma_0 = 15,000$, $\tau_1 = \tau_2 = 0.8$로 하였다. 성능 기준은 4 장의 영화 프레임이 5 장의 프레임으로 변했다고 보고, 5 장이 다시 4 장으로 올바르게 돌아갈 수 있도록 검출하였거나, 프레임을 잃어버려서 복원할 수 없더라도 어떤 타입의 프레임인지 파악하였다면 성공으로 보고, 이외의 경우는 실패로 보았다. 600 프레임의 테스트 영상이라면, 600/5 = 120 블록이 있고, 각각의 영상 400~600 장에 실험한 결과를 백분율로 표시하여 표 5에 정리하였다.

비교를 위한 기존의 방법[1]에서는 1, 2 처럼 화질이 좋지 않은 VTR에서 캡쳐한 화면과 3, 4, 5 처럼 비교적 깨끗한 화질의 화면에 모두 적용하는 문턱치를 잡기 어렵다. 실제로 4번 영상에서 최적화된 문턱치를 2번 영상에 적용시켰을 때, 30 % 남짓의 성공률을 보일 뿐이다. 따라서 실험에서는 고화질용 문턱치와 저화질용 문턱치 두 개를 사용하였다. 4 에서 최적화된 문턱치는 3, 4, 5 영상에 적용하였고, 2 에서 최적화된 문턱치는 1, 2 에 사용하였다. 전체적으로 보아 화질이 좋은 영상에서는 어느 정도 결과를 내지만, 화질이 나쁜 영상에서는 의미 있는 결과를 내지 못하고 있다. 그 이유는 비디오에서 나오는 영상은 잡음으로 인해 얼룩이 발생하고 이 때문에 움직임 벡터가 움직임을 제대로 반영하지 못하기 때문이다. 이런 현상은 동작이 강하게 나타날 때는 별 영향을 미치지 않지만, 동작이 비교적 작은 경우에는 움직임 벡터를 사용한 방법이 제대로 동작할 수 없도록 만든다.

제안된 방법에서 에러가 발생한 상황의 반 이상은 {1,1,0,0,0}을 {1,0,0,0,0}으로 인식하는 문제이다. 만약 {1,1,0,0,0}의 결과가 나왔다면, 첫 번째 프레임의 홀수줄과 두 번째 프레임의 짝수줄을 섞어서 새로운 프레임을 만들어 주고, 세 번째, 네 번째, 다섯 번째 프레임을 그대로 유지한다. 반면 {1,0,0,0,0}이 나온 경우는 첫 번째 프레임만 버리고 나머지는 그대로 내보낸다. 일반적인 경우에는 전자의 방법이 올바른 복원 방법이다. 하지만 {1,1,0,0,0}을 {1,0,0,0,0}으로 잘못 인식한 경우에서는 상황이 일반적인 경우와 약간 다르다. {1,1,0,0,0}을 {1,0,0,0,0}으로 인식했다는 것은 첫 번째 프레임만 중간 정도의 주파수 세기에서 이탈되었다는 뜻이다. 즉, 첫 번째 프레임을 제외하고는 어색한 프레임이 없다는 뜻이다. 가장 흔한 경우는 움직이는 물체가 첫 번째 프레임과 두 번째 프레임 사이에서 화면 밖으로 나가는 것이다. 이런 상황에서는 후자의 방법으로 복원하는 것과 전자의 방법으로 복원하는 것이 시각적으로 유사한 결과를 보인다.

Ⅳ. 결 론

본 논문에서는 적은 계산량으로 높은 정확도를 보이는 인버스 텔레시네 기법에 대해 소개하였다. 이 방법은 다른 시간에 촬영된 두 프레임이 홀수줄과 짝수줄로 합쳐지면서 발생하게 되는 불일치를 주파수 영역에서 감지하고, 이 정보와 전후 관계를 고려하여 동일한 것으로 추정되는 필드를 찾아낸다. 그리고 그 필드가 실제로 일치하는 지를 판단함으로써 인버스 텔레시네해도 좋은 가를 판단하는 방법이다.

이 방법은 우선 전후 관계를 고려하기 때문에 에러률이

표 5. 실험 결과 (성공률, %)
Table 5. Experimental Results (Successful results, %)

번호	기존 방법 [1]	제안된 방법
1	70.0	95.8
2	60 미만	99.2
3	100.0	100.0
4	98.3	100.0
5	85.7	100.0

낮을 뿐만 아니라, 주로 동작이 작은 부분에서 에러를 내므로, 에러 자체도 비교적 큰 문제를 일으키지 않는다. 또한 기존 방법이 고화질용 영상에만 적합한 기법인데 반해, 본 방법은 화질에 관계없이 인버스 텔레시네를 잘 수행한다. 그리고 최근에 와서 많이 등장하는 혼합형도 제대로 감지하여 인버스 텔레시네로 인한 화질 열화를 최소화 한다.

제시된 인버스 텔레시네 방법을 통하여 TV로 방영되거나 VTR로 녹화되어 있는 비디오를 MPEG으로 압축하면 일반적으로 60 필드 모두를 압축하는 것에 비하여 약 20%의 비트를 절약할 수 있다.

참 고 문 헌

[1] Ming-Chang Liu, Tsukagoshi, I., Kutner, M.A., "Real-time MPEG video encoder with embedded scene change detection and telecine inverse", Consumer Electronics, 2002. ICCE. 2002 Digest of Technical Papers. International Conference on , 2002 Page(s): 246 -247

[2] Schutten, R.J.; de Haan, G, "Real-time 2-3 pull-down elimination applying motion estimation/compensation in a programmable device", Consumer Electronics, IEEE Transactions on , Volume: 44 Issue: 3 , Aug 1998 Page(s): 930 -938

[3] Hilman, K., Hyun Wook Park, Yongmin Kim, "Using motion-compensated frame-rate conversion for the correction of 3:2 pulldown artifacts in video sequences", Circuits and Systems for Video Technology, IEEE Transactions on , Volume: 10 Issue: 6 , Sep 2000 Page(s): 869 -877

[4] 杉原 源興, 인버스 텔레시네 変換装置 , 일본 특허 P2000-217084, 2000년 8월

[5] Alan V. Oppenheim, Ronald W. Schafer, "Discreet-time Signal Processing", PRENTICE-HALL, pp 559-561

[6] "MPEG-2 Test Model 5", ISO/IEC/JTC1/Sc29/WG11/93-400, Apr. 1993

저 자 소 개

구 형 일

- 2002년 : 서울대학교 전기공학부, 학사
- 2002년 3월 ~ 현재 : 서울대학교 전기공학부 석사과정
- 주관심분야 : 영상 처리, 영상 검색

조 남 익

- 1986년 : 서울대학교 제어계측학과, 학사
- 1988년 : 서울대학교 제어계측학과, 석사
- 1992년 : 서울대학교 제어계측학과, 박사
- 1991년 ~ 1994년 : 제어계측신기술연구센터 연구원
- 1994년 ~ 1998년 : 서울시립대학교 전자공학부, 조교수
- 1999년 ~ 2001 현재 : 서울대학교 전기공학부, 조교수
- 2001-현재 부교수
- 주관심분야 : 신호 처리, 영상 처리, 적응 필터

이 종 원

- 1990.3 ~ 1994.2 한양대학교 공과대학 전자통신공학과 학사
- 1994.3 ~ 1996.2 한국과학기술원(KAIST) 전기및전자공학과 석사
- 1996.3 ~ 2002.2 한국과학기술원(KAIST) 전기및전자공학과 박사
- 2001.3 ~ 현재 삼성전자 DS총괄 System LSI 사업부 재직중,
- MPEG 및 DVD 관련 알고리듬과 Video pre/post processing 알고리듬 개발 및 하드웨어 architecture 설계 구현
- 주관심 분야 : 동영상 압축 및 영상 신호 처리가 주 관심 분야