

영상 프레임간 차등 신호에 대한 변환 부호화 방식의 비교

권계철, 권상근, 김재균
한국과학기술원 전기 및 전자

Comparison of Transforms for Image Frame Difference

Jae-Cheol Kwon, Sang-Keun Kwon, Jae-Kyoon Kim
Dept. of Electrical Engineering, KAIST

Abstract

The statistical properties of the simple interframe prediction errors, video frame difference, are derived and compared. Various transforms are applied to the prediction error to find the efficiency of the transforms. It is shown that the DCT, followed by HCT, is closest to KLT even in the prediction error domain. The HCT is found to be effective in hybrid codec for its simplicity and good performance.

I 서 론

Sub T1 급 전송속도를 갖는 영상 부호화 방식은 Inter-frame DPCM과 변환 방식 또는 VQ(Vector quantizer)를 혼합한 복합 부호화방식 (Hybrid coding) 이 일반적으로 많이 사용된다. Interframe DPCM 은 시간 방향(temporal direction)으로의 상관 관계를, 변환 부호화 방식 또는 VQ 는 시간 방향으로의 처리 후에 잔존하는 공간 영역의 상관 관계까지도 재기하여 제한된 전송률하에서 최대의 효율을 얻기 위한 것이다.[1][2][3]

현재 여러 복합 부호화기에서 채택하고 있는 Inter-frame DPCM은 크게 이동 보상을 하지 않고 예측 오차만을 전송하는 조건부 보충 방식과 이동을 보상하여 이동 벡터와 예측 오차를 전송하는 이동 보상 방식 (Motion compensated coding)으로 구별되며 변환 방식은 준 최적 변환인 DCT(Discrete Cosine Transform)가 주로 사용 된다.[4][5]

본 연구에서는 원영상에 대한 시간 영역의 정보 처리로 인해 변화하는 예측 오차의 통계적 성질을, 이동 보상 방식을 사용하지 않은 경우와 사용한 경우에 대해 각각 구하고 이들의 성질을 비교한 후 여러 직교 변환을 예측 오차에 적용했을 때 각 변환 방식의 성능을 비교하여 예측 오차 영역에서 KLT에 근접하는 성능을 가지는 변환과 우수한 성능을 가지면서 hardware구현이 간단한 변환 방식에 관해 고찰하고자 한다.

II 본 론

1. FD 및 DFD의 통계 특성

이동 보상 방식을 사용하지 않은 경우 연속된 두 frame 사이에 물체의 움직임이 존재하지 않거나 이동 보상을 사용한 경우 완전한 이동 보상이 되었다면 그들의 예측 오차인 FD(Frame Difference)나 DFD(Displaced Frame Difference)는 영상 내에 존재하는 잡음 신호의 차와 같기 때문에 그들의 통계 특성은 잡음의 통계 특성과 같게 된다. 그러나 움직임이 존재하거나 불완전한 이동 보상으로 인해 발생 되는 정보들에 의해 FD와 DFD의 통계 특성은 변화하게 된다. 여기에서는 연속된 frame 사이에서 물체가 움직이는 경우 그들의 차등 신호(FD,DFD)가 가지는 통계적인 성질을 살펴 보고자 한다.

이 FD와 DFD의 통계적 성질은 영상 자체와 관계하는 변수들, 예를 들면 영상의 종류, 움직임의 형태, 이동량 뿐 만 아니라 이동 보상을 위해 사용된 추정 방식, cost 함수 등에 따라서도 그 특성이 다소 변화하기 때문에 두 신호에 대한 구체적인 특성을 얻기는 매우 어렵다. 따라서 여기서는 균일한 배경을 가지고 있어 uncovered background에 의한 영향을 무시할 수 있는, 고체(rigid body)의 평형 이동(constant translation) 이 발생한 경우로 제한하여 그들이 가지는 일반적인 성질을 찾아 내고자 한다. 편의상 영상 데이터는 zero mean을 가지고 stationary 하며 separable한 공분산을 갖는다고 가정한다.[6]

1-1. FD 의 통계 특성

[그림1]에서 (j-1)T frame 내의 물체가 jT frame 으로 $D=(dx, dy)$ 만큼 이동했다면 위치 좌표 $p=(x, y)$ 에서의 frame 차등 신호는 원영상의 화소 밝기를 $s(\cdot)$ 라고 할 경우 식(1)과 같이 표현된다.

$$FD(p, jT) = s(p, jT) - s(p + D, jT) \quad (1)$$

where T : frame 주기

식(1)을 이용하여 FD의 통계 특성을 구하면 다음과 같다.

$$i) \text{ 분산 : } E\{[FD(\cdot) - \overline{FD(\cdot)}]^2\} = \sigma_s^2 [1 - R_s(\underline{D})] \quad (2)$$

$$ii) \text{ 공분산 : } E\{[FD(\underline{p}jT) \cdot FD(\underline{p} + \tau jT)]\} \\ = \sigma_s^2 [2R_s(\tau) - R_s(\tau + \underline{D}) - R_s(\tau - \underline{D})] \quad (3)$$

최소 거리 (lag) τ 에서의 상관 계수 $\gamma_{FD}(\tau)$ 는 공분산을 분산으로 정규화시켰을 때 식 (4)가 성립한다.

$$\gamma_{FD}(\tau) = \frac{[2R_s(\tau) - R_s(\tau + \underline{D}) - R_s(\tau - \underline{D})]}{2[1 - R_s(\underline{D})]} \quad (4)$$

1-2. DFD의 통계 특성

[그림1]에서 이동 보상 방식을 사용한 경우 정확한 이동 보상이 되었다면 식 (5)와 같은 관계가 성립하며 위치 p 에서의 DFD($\underline{p}jT$)는 식 (6)와 같다.

$$s(\underline{p}jT) = s(\underline{p} - \underline{D}, (j-1)T) \quad (5)$$

$$DFD(\underline{p}jT) = s(\underline{p}jT) - s(\underline{p} - \underline{D}, (j-1)T) \quad (6)$$

그러나 실제 부호화 방식에서 사용되는 이동 보상 방식은 정확한 이동 벡터 \underline{D} 를 구할 수가 없다. $\hat{\underline{D}}$ 를 실제 부호화기에서 사용하는 이동 추정 방식을 이용하여 추정된 이동 벡터라고 할 때 실제 DFD(\cdot)는 식 (7)과 같이 표현될 수 있다.[7]

$$DFD(\underline{p}jT) = s(\underline{p}jT) - s(\underline{p} - \hat{\underline{D}}, (j-1)T) \quad (7)$$

식 (7)을 이용하여 DFD의 여러 통계적 성질을 구할 때 다음과 같다.

$$i) \text{ 분산 : } \sigma_{DFD}^2 = E\{[DFD(\cdot) - \overline{DFD(\cdot)}]^2\} \\ = \sigma_s^2 [1 - R_s(\Delta \underline{D})] \quad (8)$$

$$ii) \text{ 공분산 : } E\{[DFD(\underline{p}jT) \cdot DFD(\underline{p} + \tau jT)]\} \\ = \sigma_s^2 [2R_s(\tau) - R_s(\tau + \Delta \underline{D}) - R_s(\tau - \Delta \underline{D})] \quad (9)$$

식(9)로부터 최소 거리 τ 에서 DFD의 상관 계수는 식 (10)과 같다.

$$\gamma_{DFD}(\tau) = \frac{2R_s(\tau) - R_s(\tau + \Delta \underline{D}) - R_s(\tau - \Delta \underline{D})}{2[1 - R_s(\Delta \underline{D})]} \quad (10)$$

여기서 $\Delta \underline{D}$ 가 확률 밀도 함수 $P_{\Delta \underline{D}}(\cdot)$ 를 가지는 random variable 이라 가정하면 i) ii)의 분산과 공분산은 각각 아래와 같이 표현된다.

$$E[\sigma_{DFD}^2] = \sigma_s^2 [1 - \int_{-\infty}^{\infty} P_{\Delta \underline{D}}(k) \cdot R_s(k) dk] \\ E[DFD(\underline{p}jT) \cdot DFD(\underline{p} + \tau jT)] \\ = \sigma_s^2 \int_{-\infty}^{\infty} [R_s(k) - R_s(\tau + k) - R_s(\tau - k)] \cdot P_{\Delta \underline{D}}(k) dk$$

3. FD와 DFD의 특성 비교

원영상의 정규화된 공분산이 $R_s(\tau) = \rho^{|\tau|}$ 로 표현된다고 가정하면[6] 위 식 (2),(4),(8),(10)으로부터 다음과 같은 일반적인 FD와 DFD의 성질을 구할 수 있다.

$$1. \gamma_{FD}(\tau) < R_s(\tau) \quad , \quad \gamma_{DFD}(\tau) < R_s(\tau).$$

2. 이동량(D)이 증가할 수록 FD의 temporal 상관 관계는 감소하고 분산과 spatial 상관 관계는 증가한다.

3. 추정 오차(ΔD)가 증가할 수록 DFD의 temporal 상관 관계는 감소하고 분산과 spatial 상관 관계는 증가한다.

이동 보상 방식을 사용한 경우 이동이 발생한 영역내에서는 모든 block에 대해 $\Delta \underline{D}$ 가 같고 $|\Delta \underline{D}| < |\underline{D}|$ 가 성립하다고 가정하면 식(2),(8)로부터 두 신호의 분산은 항상 식(11)과 같은 관계가 성립한다.

$$\sigma_{FD}^2 - \sigma_{DFD}^2 > 0 \quad (11)$$

두 신호의 상관 관계를 비교하기 위해 각 신호의 1-Dim 상관 관계의 절대치 차를 식(12)이라 할 때 아래와 같은 관계를 갖는다.

$$f(\tau_x) = |\gamma_{FD}(\tau_x)| - |\gamma_{DFD}(\tau_x)| \quad (12)$$

$$f(\tau_x) \leq 0, \quad \text{for } \tau_1 < \tau_x \leq \tau_2$$

$$f(\tau_x) > 0, \quad \text{for elsewhere}$$

$$\tau_1 = \frac{1}{2 \ln \rho} \ln \left[\frac{\rho^{\tau_1} (1 - \rho^{\Delta D x})}{4 - 3\rho^{Dx} - 3\rho^{\Delta D x} - \rho^{-\Delta D x} + 2\rho^{Dx + \Delta D x} + \rho^{Dx - \Delta D x}} \right] \\ \tau_2 = \frac{1}{2 \ln \rho} \ln \left[\frac{\rho^{Dx} (1 - \rho^{\Delta D x})}{\rho^{Dx} (1 - \rho^{-\Delta D x}) + \rho^{-\Delta D x} - \rho^{\Delta D x}} \right]$$

위의 결과로부터, \dots 의 상관 관계는 DFD보다 큼을 알 수 있다. [그림2]는 두 신호의 상관 관계를 나타낸다.

2. 예측 오차 영역에서의 변환 방식

일반적으로 어떤 신호의 상관 관계는 그 신호가 가지는 각 주파수 성분의 에너지 분포와 관계하는데 상관 관계가 작다는 것은 영상내에 복잡한 부분이 많고 변환 영역에서는 높은 주파수 성분의 계수들이 가지는 에너지가 상대적으로 증가한다는 것으로 이는 에너지 집중 효과에 의존하는 변환 부호화 방식의 효율이 적다는 것을 의미한다. 따라서 상관 관계가 작은 FD나 DFD의 변환 부호화 효율은 원영상의 경우 보다는 크게 떨어지지만 영상 전화 부호화기와 같은 낮은 전송률을 가지는 시스템의 부호화 방식에서 원하는 품질로 정보를 전송 하기 위해서는 이들 신호의 양자화 만으로는 어렵기때문에 약간 남아 있는 spatial 상관 관계 까지도 제거하는 변환 부호화 과정을 거쳐야 한다.

변환 부호화 방식에는 여러가지가 있으나 여기서는 대표적인 내가치의 직교 변환 방식에 대해 이들이 예측 오차 영역에서 가지는 성능을 고찰하고자 한다. 비교 검토될 변환 부호화 방식은 신호의 통계 특성에 의존하는 최적 변환인 KLT(Karhunen-Loeve Transform)와 stationary 하고 상관 계수가 큰 지수형 공분산을 갖는 1차 Markov process에 대해 KLT에 가장 근접하는 성능을 나타내는 것으로 알려져 있는 DCT, 비정형 변환으로 곱셈 연산이 없고 매우 간단하게 구현할 수 있는 WHT(Walsh-Hadamard

Transform), WHT의 변형체로 WHT정도의 단순성을 가지면서 DCT에 가까운 성능을 가지는 HCT(High Correlation Transform)이다.

원영상과 그의 차등 신호는 통계 특성이 차이가 있기 때문에 이들을 변환 부호화 했을 때 부호화 성능도 달라지게 된다. 1.에서 얻어진 결과를 바탕으로 식(13)과 같은 정규화된 공분산 행렬을 갖는 원영상에 대해 변환 부호화의 성능을 고찰하고자 한다.[6]

$$COV(X) = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix} \quad (13)$$

사용될 변환 부호화 성능의 평가로는 아래 식(14)로 정의되는 에너지 집중량으로 한다.

$$\eta_E(M) = \frac{\sum_{i=1}^M \sum_{j=i}^M |y_{ij}|}{\sum_{i=1}^N \sum_{j=i}^N |y_{ij}|} \times 100 [\%] \quad (14)$$

여기서 y_{ij} : 변환영역 공분산 행렬의 ij 번째 성분

N : Blocksize

M : retained coefficient number

식(13)에서 $N=8$ 일 때 상관 계수 ρ 를 변화시키에 따라 변화하는 여러 변환들의 에너지 집중량 $\eta_E(M=3)$ 이 [그림3]에 나타나 있다. 이 그림으로부터 원영상에 대한 변환 효율은 상관 계수의 크기에 관계없이 DCT가 KLT에 가장 근접함을 볼 수 있으며 HCT변환이 DCT에 근접하며 WHT 변환은 극단적인 상관 계수 ($\rho \neq 0, \rho \neq 1$)에서만 제외하고는 위 변환 방식과는 다소 성능 차이가 있음을 볼 수 있다.

다음은 예측 오차 영역에서 각 변환 방식의 성능을 비교하고자 한다. 이등 보상 방식을 사용하지 않은 경우와 사용했을 경우 각각의 예측 오차의 상관 관계를 나타내는 식(4),(10)은 같은 형태를 가지고 있으므로 두 식을 아래 식(15)과 같이 일반화시켜 사용하도록 한다.

$$\gamma_e(k) = \frac{2R_s(k) - R_s(k+Q) - R_s(k-Q)}{2(1 - R_s(Q))} \quad (15)$$

여기서 $k=(k_x, k_y)$: 화소간거리(lag)

$$\underline{Q} = \begin{cases} \underline{D} [= (D_x, D_y)] : \text{in FD} \\ \underline{\Delta D} [= (\Delta D_x, \Delta D_y)] : \text{in DFD} \end{cases}$$

원영상의 정규화된 공분산 함수가 $R_s(k) = \rho^{|k_x|} \cdot \rho^{|k_y|}$ 이라고 하고 $\gamma_e(\cdot)$ 의 1-Dim 상관 관계만 고려한다면 공분산 함수는 식(16)로 표현되며 이로부터 얻어지는 공분산 행렬 COV(E)는 식(17)와 같다.

$$\gamma_e(k_x) = \frac{2\rho^{|k_x|} - \rho^{|k_x-Q_x|} - \rho^{|k_x+Q_x|}}{2(1-\rho^{|Q_x|})} \quad (16)$$

$$COV(E) = \begin{bmatrix} \gamma_e(0) & \gamma_e(1) & \gamma_e(2) & \dots & \gamma_e(N-1) \\ \gamma_e(1) & \gamma_e(0) & \gamma_e(1) & \dots & \gamma_e(N-2) \\ \gamma_e(2) & \gamma_e(1) & \gamma_e(0) & \dots & \gamma_e(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_e(N-1) & \gamma_e(N-2) & \gamma_e(N-3) & \dots & \gamma_e(0) \end{bmatrix} \quad (17)$$

원영상의 상관 계수가 $\rho = 0.95$ 일 때 Q_x 의 크기에 따른 에너지 집중량의 변화 형태를 보면 [그림4]과 같다.

이로부터 Q_x 가 극히 작은 경우에는 예측 오차 신호는 거의 random 잡음과 같은 성질을 가지게 될 것이므로 모든 변환 방식의 성능이 비슷하나 Q_x 가 증가함에 따라 예측 오차 신호가 가지는 상관 관계는 커지게 되므로 각 변환의 성능이 급격히 증가하는 것을 볼 수 있으며 예측 오차 영역에서도 DCT는 다른 직교 변환보다도 KLT에 가장 근접하며 HCT 변환은 DCT와 WHT 중간 정도의 성능을 가진다.

지금 까지 원영상과 예측 오차 영역에서 여러 변환 방식들의 성능을 제한된 가정의 범위 내에서 비교하여 보았다. 얻어진 결과들이 실제 영상에서도 그대로 적용되는지를 확인하기 위해 실제 영상의 예측 오차들을 사용하여 여러 변환들의 성능을 살펴보았다. 모의 실험에 사용된 영상은 window size가 288×352 이며 8 bit 해상도를 갖는 Miss America 영상이다.

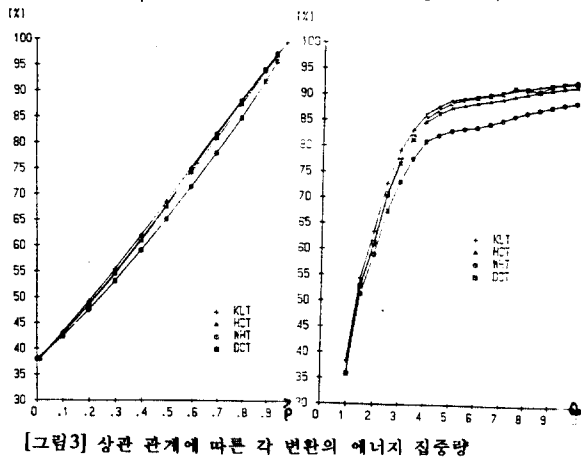
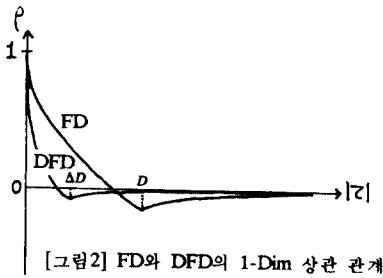
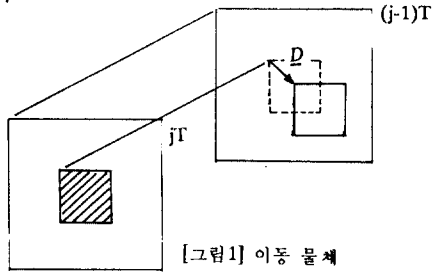
[그림5]와 [그림6]은 이 영상에 대해 이등 보상 방식을 사용하지 않은 경우와 사용한 경우의 예측 오차를 2D 변환한 후 계수들을 zigzag scan 했을 때 처음 10개의 변환 계수가 가지는 상대적인 에너지양(%)을 나타낸다. 여기서 이등 탐색 방식은 3 step search 방법을, cost 함수는 MAE(the Mean of the Absolute Error) 를 사용하였으며 block size는 8×8 이다.[8] 이 그림들로부터 이등 보상 방식을 사용하지 않은 경우가 변환 효율은 더 좋으나 움직임의 양에 따라 급격한 성능 변화를 나타냄을 볼 수 있고 이등 보상 방식을 사용한 경우에는 이동량에 관계없이 비교적 완만한 성능 변화를 나타내고 있다. 또 실제 영상에 대한 각 변환의 효율을 보면 DCT가 KLT에 가장 근접하는 성능을 보이며 경우에 따라서는 더 우수한 성능을 나타내기도 한다. 이로부터 실제 영상의 예측 오차 영역에서도 DCT는 hardware구현이 가능한 직교 변환중 최고의 성능을 나타내며 HCT는 DCT와 WHT사이의 성능을 가진다.

그런데 실제 다른 여러가지 scheme들과 결합되어 동작하는 복합 부호화기에서는 변환 부호화의 적은 성능의 차가 전체 시스템의 성능에 큰 영향을 미치지 않는 않기 때문에 DCT의 사용만이 최선이라고 할 수는 없다. 그리고 DCT는 아직까지는 hardware구현이 복잡하고 고정 소수점 연산으로 구현할 경우 필연적으로 발생하는 연산 오차로 인해 DCT처리기 자체만의 결함을 가지는 단점이 있어 복합 부호화 방식에서는, hardware는 WHT정도의 복잡도를

가지면서 비교적 좋은 성능을 나타내는 HCT 의 사용이 여러 관점에서 더 유용할 수 있다.

III 결

지금까지 복합 부호화 방식에서 사용되는 Interframe DPCM 의 두 가지 방식에 대한 예측 오차 신호 (FD,DFD)들의 통계적인 성질을 고찰 하였으며 이들 신호에 여러 변환 방식을 적용시켰을 때 각 변환 방식들의 성능을 비교하였다. 이론에서 얻어진 것과 같이 변환 성능만이 고려되는 경우 이동 보상 방식을 사용하지 않은 경우가 더 좋은 결과를 주며 여러 변환 방식중 DCT가 예측 오차 영역에 서도 KLT에 가장 근접하는 성능을 보이며 HCT도 비교적 우수한 결과를 보여 준다. 그중 HCT는 hardware구현시 DCT보다도 동작 속도가 빠르고 간단하게 구현할 수 있어 실시간 처리에 적합하며 실제 64Kbps 영상 전화 부호화기에 변환 방식으로 DCT와 HCT를 적용해본 결과 성능차는 거의 없는 것으로 나타나 HCT의 유용성이 확인되었다.



[REFERENCE]

- [1] J.R. Jain, A.K.Jain "Displacement measurement and it's application in interframe image coding" IEEE Trans.on comm COM-29 1981
- [2] 김 재균 의 9인, "영상 정보의 부호화 기술 개발 연구" 최종 보고서, 한국 과학 기술원, 과학 기술처, 1987
- [3] K.Mutsuda, et al, "Motion compensated 384Kb/s video codec for Teleconference" GLOBECOM '85, Dec 1985
- [4] Robert D. Widergren, "Full color Motion video at 56Kb/s ; Military and civil applic." GLOBECOM '85, Dec 1985
- [5] NEC Corporation, "NETEC-XD; 56KB Motion codec for International and Domestic Motion video Teleconference"
- [6] R.J. Clark, "Transform Coding of Image", Academic Press, 1985
- [7] 권 상근, "이동 보상형 복합 부호화기를 위한 Block별 frame 차등 신호의 추정 방법에 관한 연구" 한국 과학 기술원 석사 학위 논문 1987
- [8] T. Koga, et al, "Motion compensated Interframe coding for video conference" Nat'l Tele Conf 1981 New orleans L.A. U.S.A.

