

C군의 네버엔딩 스토리, 디지털 영상처리의 이해 - 6

TV를 보다 보면 화면이 일반 영상으로만 채워지는 적이 거의 없습니다. 항상 방송국 로고가 화면의 상단 어딘가에 있고, 뉴스와 예능 등 대다수의 TV 프로그램들에는 자막이 쉴 새 없이 나옵니다. 이렇게 영상 위에 로고나 문자 등의 그래픽이 끊임없이 흐르는 것이 너무도 익숙하고 자연스러운 나머지, 어떤 원리로 이런 것들이 가능한지 궁금하지조차 않은 것이 사실입니다. 그러나 이런 당연하고 자연스러운 것들을 들추어 그 원리를 설명하는 것이 본 연재가 살아가는 길이므로, 궁금한 사람 한 명 없다 해도 오늘도 묵묵히 그 원리를 구구절절 물어보겠습니다.



그림 1. 영상과 문자 그래픽의 합성 예

[그림 1]은 저작권 관련 이슈가 발생하지 않도록 '방송과 기술' 홈페이지에서 이미지를 다운로드 한 뒤, 그 위에 '방송과 기술'이라는 문자 그래픽을 얹은 것입니다. [그림 1]의 배경 영상이 정지영상이 아니라 동영상이라면 일반적인 방송영상과 다를 것이 없습니다. 그러면 간단한 산수를 이용하여 [그림 1]의 배경영상과 문자 그래픽 합성원리를 설명한 후, 실제 방송제작 시스템에서 어떻게 처리되는지를 간략히 설명하겠습니다.



그림 2. 배경영상(좌)과 문자 그래픽(우)

[그림 1]은 [그림 2]의 배경영상(좌)에 문자 그래픽(우)을 합성한 것입니다. 그런데 아무 지능이 없는 기계로 [그림 2]의 배경영상과 문자 그래픽을 합성하려면 어떻게 해야 할까요? 크로마키처럼 문자 그래픽의 검은색 부분만 배경영상으로 대체하면 [그림 1]처럼 합성이 될까요? 만약 문자 그래픽의 검은색 부분만 배경영상으로 대체하는 방법을 쓴다면, 문자 그래픽의 문자 색상이 검은색일 때는 어떻게 해야 할까요? 아주 약간 구체적으로 생각해 보았을 뿐인데 왠지 [그림 2]의 배경영상과 문자 그래픽을 합성하는 것이 간단치 않아 보인

다. 사실 문자의 색상에 상관없이 [그림 2]의 배경영상과 문자 그래픽을 합성하려면 [그림 3]과 같이 문자 그래픽의 Fill과 Key 두 개의 영상이 필요합니다.



그림 3. 문자 그래픽의 Fill(좌) 영상과 Key(우) 영상

[그림 3]의 좌측은 문자 그래픽의 Fill 영상이고 우측은 Key 영상입니다. Fill 영상은 배경영상에 합성될 그래픽 자체라고 보시면 되고 화소마다 R, G, B 값을 갖는 컬러영상이 일반적입니다(디지털 영상의 각 화소에서 R, G, B 값 등이 가지는 의미는 이전 연재에 설명되어 있으니 참고하시기 바랍니다). Key 영상은 각 화소가 배경영상과 Fill 영상 중 어느 것을 선택해야 하는지 나타낸 것이라고 보면 되며 화소마다 밝기 값만을 가진 흑백영상입니다. [그림 4]는 [그림 1]의 합성된 영상을 만들기 위한 연산의 개념을 나타낸 것입니다.



그림 4. 배경영상과 문자 그래픽 합성을 위한 연산 개념도

[그림 4]에서 Key 영상의 밝은 부분은 Fill을 통과시키고, 어두운 부분은 Fill을 막는 마스크 같은 것으로 생각하면 됩니다. 배경영상 입장에서 Key 영상이 Fill 영상을 통과시키는 부분은 배경영상을 막는 것과 같은 효과가 있고, Key 영상이 Fill 영상을 막는 부분은 배경영상을 통과시키는 것과 같은 효과가 있으므로 Key의 흑백 반전영상을 마스크로 사용하는 것과 같은 효과가 있는 것으로 해석하면 됩니다. 이와 같이 Key 영상과 Key 흑백 반전영상을 각각 Fill 영상과 배경영상의 마스크로 해석하여 각 화소의 R, G, B 값을 이용한 식으로 정리하면 [수식 1]과 같습니다.

[수식 1]에서 사용된 모든 R , G , B , Key 등의 값들은 각각의 영상에서 같은 위치의 화소들이 갖는 값들이며 R_{Output} , G_{Output} , B_{Output} 은 [그림 4]의 결과영상에서 각 화소가 갖는 R , G , B 값, R_{Fill} , G_{Fill} , B_{Fill} 는 [그림 4]의 문자 그래픽의 Fill 영상에서 각 화소가 갖는 R , G , B 값, $R_{Background}$, $G_{Background}$, $B_{Background}$ 는 [그림 4]의 배경영상에서 각 화소가 갖는 R , G , B 값, Key 는 [그림 4]의 Key 영상에서 각 화소가 갖는 밝기 값을 나타냅니다. [수식 1]에서 $\frac{Key}{255}$ 를 사용한 이유는 0~255 사이의 값을 갖는 8비트 디지털 영상을 사용한다는 가정 하에 Key 영상의 화소값을 0~1 사이의 값으로 조절하여 마스크(차단→통과)와 같은 역할을 하도록 조절하기 위한 것입니다. 만약, 10비트 디지털 영상을 사용한다면 Key 영상의 화소들이 가지는 밝기 값의 범위가 0~255가 아닌 0~1023이 되므로 $\frac{Key}{255}$ 를 $\frac{Key}{1023}$ 로 바꾸어야 합니다. [수식 1]에서 쉽게 알 수 있듯이 Key 영상에서 255의 값을 갖는 화소의 위치에 대응하는 R_{Output} , G_{Output} , B_{Output} 는 R_{Fill} , G_{Fill} , B_{Fill} 과 동일하며, Key

$$R_{Output} = R_{Fill} \times \frac{Key}{255} + R_{Background} \times \left(1 - \frac{Key}{255}\right)$$

$$G_{Output} = G_{Fill} \times \frac{Key}{255} + G_{Background} \times \left(1 - \frac{Key}{255}\right)$$

$$B_{Output} = B_{Fill} \times \frac{Key}{255} + B_{Background} \times \left(1 - \frac{Key}{255}\right)$$

수식 1. 배경영상과 문자 그래픽 영상의 합성을 위한 산술식

영상에서 0의 값을 갖는 화소의 위치에 대응하는 R_{Output} , G_{Output} , B_{Output} 는 $R_{Background}$, $G_{Background}$, $B_{Background}$ 와 동일하게 됩니다. 다시 말하면, Key 영상이 흰색인 부분은 Fill 영상과 같고 검은색인 부분은 배경영상과 같은 결과영상을 [수식 1]을 통해 얻을 수 있게 됩니다. 지금까지의 설명만으로 Fill과 Key 영상을 이용한 영상합성에 관한 전반적인 이해에 충분할 것 같습니다. 그렇다면 여기에서 C군이 질문을 하나 해보겠습니다. 결과영상의 각 화소의 위치에 대응하는 Fill 영상과 배경영상의 화소값을 선별적으로 선택하는 것이 Key 영상의 목적이라면, 왜 0과 1의 두 값만 사용하지 않고 복잡하게 0~255(8비트 영상의 경우)까지의 값을 사용하게 하였을까요? 이에 관한 대답은 크게 두 가지가 있을 것 같습니다. 첫 번째 이유는, [그림 5]와 같이 Key 영상을 확대해보면 흰색과 검은색의 경계가 사선으로 존재할 경우 이 경계를 따라 흰색과 검은색의 중간에 해당하는 회색 화소들이 존재하는 것이 보입니다. 이 회색 화소들은 [수식 1]에 의해 Fill 영상과 배경영상이 적당히 섞이도록 하는 역할을 수행합니다. Key 영상의 사선 경계를 따라 Fill 영상과 배경영상이 적당히 섞이게 될 경우 결과영상에서 Fill 영상과 배경영상이 만나는 경계는 매우 부드러운 경계를 형성하여 영상의 품질이 좋아 보입니다. 하지만 [그림 6]과 같이 모든 화소에서 0 또는 1의 흑과 백의 값만을 갖는 Key 영상을 사용하게 되면 결과영상에서 Fill 영상과 배경영상이 만나는 경계가 [그림 6]의 Key 영상과 같이 매우 거칠어져서 영상의 품질이 떨어져 보입니다.

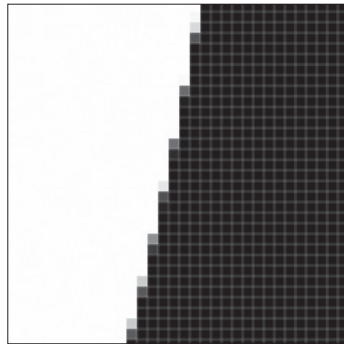


그림 5. 0~255의 값을 갖는 Key 영상

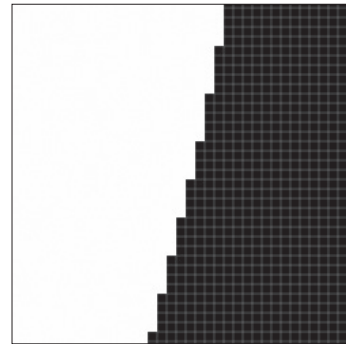


그림 6. 0 또는 1의 값을 갖는 Key 영상

두 번째 이유는 [그림 7]처럼 Fill 영상을 반투명으로 합성하는 경우가 많습니다. 이렇게 반투명으로 Fill 영상과 배경영상을 합성할 경우의 Fill과 Key 영상은 [그림 8]과 같습니다. 불투명으로 Fill 영상을 합성하는 [그림 3]의 예와 비교하여 [그림 8]의 Key 영상은 문자 부분이 회색인 것을 알 수 있습니다. 이 회색 부분을 [수식 1]에 대입하면 Fill 영상과 배경영상이 적당히 섞이게 됩니다. Fill 영상과 배경영상이 적당히 섞여서 Fill 영상의 문자 부분에 배경영상이 비치는 것 같이 보이므로 반투명 같은 효과가 나겠죠?



그림 7. 문자 그래픽의 반투명 합성 예



그림 8. 문자 그래픽의 반투명 합성을 위한 Fill과 Key 영상 예

지금까지 Fill과 Key 영상을 이용한 영상합성의 원리를 알아보았습니다. 그렇다면 이런 원리가 방송제작 시스템에서는 어떻게 구현되는 것일까요? [그림 9]는 방송제작 시스템에서 영상과 문자 그래픽을 합성하는 부분만을 간략하게 나타낸 것입니다. 카메라 또는 비디오 서버 등에서 영상이 출력되고, 문자 발생기 또는 기타 그래픽 장비를 통하여 영상에 얹히게 되는 그래픽의 Fill과 Key 영상이 출력됩니다. 이 모든 영상출력들은 스위처라는 장비에 입력되고, 스위처에서 [수식 1]과 같은 연산을 통하여 영상과 문자 그래픽 등이 합성된 결과영상을 만들어 냅니다.



그림 9. 방송제작 시스템에서의 영상합성

지금까지의 설명으로 별로 어렵지 않게 영상합성에 관해 이해가 되었을 것 같습니다. 다음 연재에서는 Fill과 Key 영상을 이용한 영상합성의 고급편이라 할 수 있는 Pre-multiplied Alpha 방식에 관해 설명하겠습니다. 📺