

이산 푸리에 변환과 Zernike 모멘트를 이용한 텍스처 영상 검색

(Texture retrieval using DFT and Zernike moments)

현대전자산업주식회사 정보통신 연구소 기반기술연구실

심 동 규 (dgsim@hei.co.kr), 김 해 광 (cerise@hei.co.kr)

【요약】

본 논문은 이산 푸리에 변환(DFT : Discrete Fourier Transform)과 Zernike 모멘트를 사용하여 텍스처 영상을 검색하는 방법에 관한 것이다. 본 논문은 DFT를 이용하여 입력 영상의 이동에 관계없이 텍스처 영상을 검색할 수 있으며, DFT의 정규화된 에너지 특성을 이용함으로써 스케일 변화에도 관계없는 텍스처 영상 검색을 가능하게 한다. 또한 이러한 정규화된 에너지 평면에 회전에 불변한 특성을 가진 Zernike 모멘트를 사용함으로써 회전에도 관계없이 텍스처 영상을 검색할 수 있다. 결과적으로 제안한 알고리즘을 통하여 인간의 시각적 인식 능력과 같은 이동, 스케일, 회전에 관계없이 텍스처를 신속하고 정확하게 검색할 수 있다. 본 논문에서는 여러 가지 실험영상을 통하여 제안한 알고리즘이 효과적임을 보였다.

I. 논문의 개요

본 논문은 텍스처 영상을 이동, 스케일, 회전, 그리고 변형에 관계없이 신속하게 검색할 수 있도록 한 텍스처 영상 검색 방법에 관한 것이다. 본 논문에서는 이산 푸리에 변환

(DFT)과 Zernike 모멘트를 이용하여 이동, 스케일 및 회전에 불변하는 텍스처 특징값을 추출함으로써 이동, 스케일 그리고 회전에 관계없이 텍스처 영상을 검색할 수 있도록 하였다. 여기서 텍스처란 물체의 질감을 나타내는 특성이다. 예를 들어, 여러 개의 나무의 표면은 서로 다른 나무이지만 인간은 이것을 모두 나무 껍질로 판단하며, 그 이유는 같은 질감을 갖고 있기 때문이다.

현재까지의 데이터 검색은 주로 문자 자체를 검색하는 문자기반 검색 (text-based search)이었다. 그러나 근래에 들어서 멀티미디어 기술의 발전으로 방대한 양의 멀티미디어 데이터가 생성됨에 따라 문자는 물론 영화, 합성영상, 정지영상, 동영상, 음성, 그리고 음악과 같은 멀티미디어 데이터 베이스 검색이 요구되고 있다. 또한 사용자가 보다 개념적인 단어로 관계된 데이터의 검색을 요구하는 경향이 커지고 있다. 이러한 이유로 멀티미디어 데이터의 검색은 점차 중요한 과학기술 분야로 커질 것이며, 이것은 인터넷이나 방송과 같은 멀티미디어 매체의 엄청난 정보의 바다에서 원하는 데이터를 신속하게 찾게 해 줄 것으로 기대된다. 특히 영상은 텍스처의 조합으로 볼 수 있어 텍스처 영상 검색은 매우 중요

한 분야중의 하나이다.

기존의 텍스처 영상 검색 기법은 영상의 통계적 특징을 이용하는 방법이 주로 이용되어 왔다. 한 화소의 밝기값은 주변값에 의하여 결정된다는 MRF (Markov Random Field)나 주변 화소들 사이의 관계를 이용하는 co-occurrence 행렬 등이 많이 이용되어 왔다. 이러한 방법이 수학적으로 효과적이나 실제적으로는 계산량이 과다하다는 단점을 가지고 있다. MRF의 경우 MRF로 모델링한 후 통계 파라미터를 구하기 위하여 통계적 반복기법을 사용하게 된다. 이러한 방법은 최적 해를 구하는 것 자체가 보장되지 않으며 많은 계산량을 필요로 한다. Co-occurrence 행렬에 의한 방법은 여러 방향에 대한 또한 각 거리에 대한 변화량을 계산하는 것으로 대단히 많은 메모리를 필요로 하게 된다. 최근에는 텍스처 영상의 검색 기법으로서, 주파수 영역을 특징으로 이용하는 방법이 많이 이용되고 있다. 여기서 필터부내의 필터(필터1 ~ 필터 N)로 Gabor나 Wavelet필터 등이 사용된다. Gabor 필터와 같은 경우 주파수 영역을 가장 효과적으로 나눌 수 있어 많이 이용되고 있으며, Wavelet 필터는 주파수 영역을 인간의 시각적 특성이 고려된 형태로 나눌 수 있어 많이 이용되고 있다. 그러나 이러한 방법들도 모두 많은 시간을 필요로 한다. 이렇게 계산량이 많은 방법은 실제로 대용량 멀티미디어 데이터 검색에는 적합하지 않다. 예를들어 데이터 베이스를 구성하는데서 너무 많은 시간과 비용을 요구할 것이며, 실제 사용자가 검색에 소요되는 시간이 길어지면 이것의 사용을 기피할 것이다.

본 논문에서는 DFT평면에서 Zernike 변환을 하여 이동, 스케일, 및 회전에 관계없이

텍스처를 검색할 수 있는 방법을 제안하였다. 제 2장에선 제안한 알고리즘을 설명하고, 3장에서 실험결과를 보인 후, 4장에서 결론을 맺는다.

II. 텍스처 검색을 위한 제안한 DFT평면에서의 Zernike 모멘트 특징

그림 1은 이동, 스케일 및 회전에 무관한 특징을 추출하기 위한 본 논문에 의한 DFT와 Zernike 모멘트를 이용한 텍스처 영상 검색 방법을 보인 흐름도이다. 그림에서 보듯이 입력 영상 평면을 주파수 평면으로 변환하고, 이 주파수 평면에서 이동에 무관한 에너지를 추출한다. 정규화된 에너지 평면을 얻기 위하여 컷-오프 주파수를 추출하고 이 값에 따라 스케일 변화에 무관한 특징을 가진 정규화된 에너지 평면으로 변환한다.

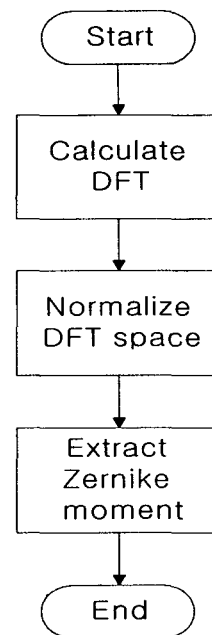


그림 1. 제안한 텍스처 특징 추출 흐름도

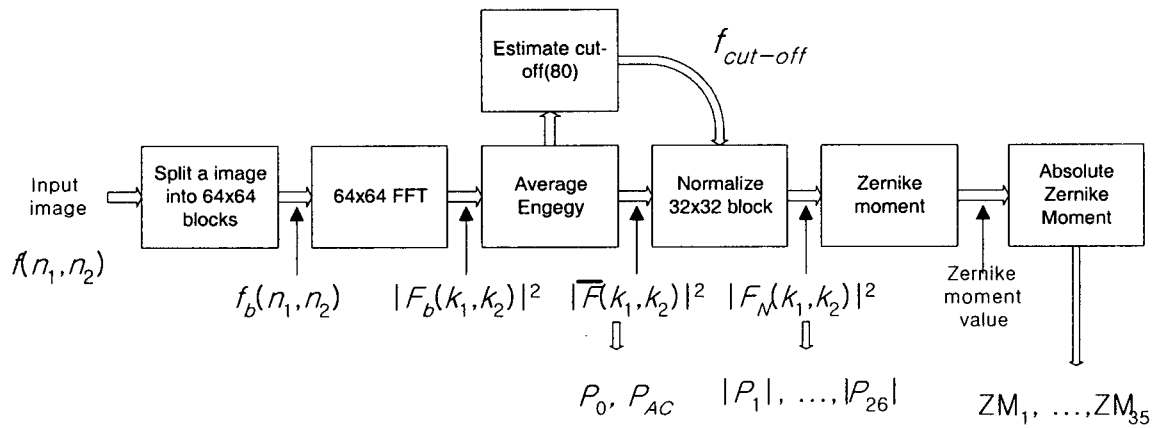


그림 2. 제안한 텍스처 특징 추출 블록도

마지막으로 정규화된 에너지 평면에서 Zernike 모멘트를 추출하여 이 값을 특징으로 사용한다. 추출한 각각의 특징과 데이터 베이스에 미리 저장된 텍스처 특징 정보를 비교하여 텍스처 영상을 결정한다.

제안한 텍스처 특징 추출은 그림 2에 도시된 바와 같이, 영상 분할기에서 입력 영상을 64*64 크기로 분할을 하고, 에너지 추출기에서 DFT를 이용하여 이동에 무관한 DFT 에너지를 추출한다. 여기서 분할된 영상을 $f_b(n_1, n_2)$ 로 표현한다. DFT는 가장 잘 알려진 변환 중의 하나로 1차원의 시간 평면을 1차원 주파수 평면으로 변환하거나, 2차원 공간 평면 신호를 2차원 주파수 평면으로 분리하는 특징을 가지고 있다. 이 DFT는 각 주파수 평면을 균일하게 나누고 각 주파수 별로 효과적으로 나눌 수 있는 변환으로 알려져 있다. DFT의 여러가지 특징이 있으나 본 발명에 사용한 이동에 무관한 특징을 추출하기 위하여 다음과 같은 특성을 이용하였다. 여기서 분할된 영상의 주파수 변환 결과를 $F_b(k_1, k_2)$ 라고 한다. 그리고 평균 에너지 추출기는 에너지 추출기

에서 얻어지는 에너지를 제공하고, 블록의 개수로 정규화한 평균 에너지를 추출하게 된다. 다음으로, 스케일 변화에 무관한 특징을 추출하기 위해서 주파수 평면을 정규화된 에너지의 평면으로 변환한다. 즉, 어떤 신호에 대한 주파수 평면은 원 신호에 대한 정보를 모두 가지고 있다. 원 신호 평면에서 텍스처 검색의 문제는 주파수 평면에서의 신호의 검색으로 볼 수 있다. 주파수 평면의 절대값 평면은 원 신호의 이동에 무관한 특징임으로 이동은 고려할 필요가 없어진다. 스케일이 변화할 경우 주파수 평면에서 확대 및 축소가 됨을 알 수 있다. 이것은 DFT의 특성에 의한 것으로 반비례하는 특징을 가지고 있다. 원 신호 $f(t)$ 에 대한 FT값이 $F(w)$ 라 할때, $f(t/s)$ 의 FT값은 $2F(2w)$ 로 표현된다. 이러한 스케일 변화 문제를 해결하기 위하여 그림 3에서 보이는 것처럼 DFT평면의 모양을 비교한다는 관점에서 원점을 중심으로 전체 에너지의 80%가 들어가는 원을 고려하여 정규화 시킨다면, 스케일 변화에 의한 주파수 평면의 확대 및 축소를 해결할 수 있다. 따라서 컷-오프기는 상기 평

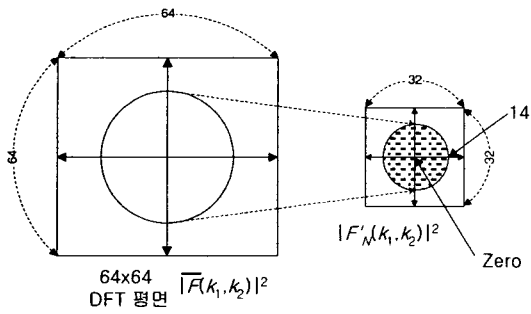


그림 3. 컷오프 주파수와 정규화된 주파수 평면

균 에너지 추출기에서 얻어지는 에너지의 80%가 되는 에너지로 컷-오프한다. 즉, 80%의 에너지가 포함되는 cut-off 주파수를 cut-off(80%)된 에너지로 표현한다. 이 cut-off(80%) 에너지에 의하여 주파수 평면을 보간이나 샘플링 기법에 의하여 정규화 시킬 수 있다. 이러한 과정을 통하여 원 영상 평면에서의 스케일 변화에 무관한 특징을 추출할 수 있다. 이렇게 구한 cut-off(80%)을 가지고 스케일 변화에 무관한 특징을 위하여 정규화기는, 정규화를 수행하여 정규화 에너지 평면을 구한다. 이 정규화된 DFT 평면을 특징값을 분할하여 P1 ~ P26까지의 26개의 특징값을 추출하여 이것을 특징 벡터로 사용한다.

한편, Zernike 모멘트 추출기는, 정규화기에서 얻어지는 정규화된 에너지 평면에서 회전에 무관한 특징을 추출한다. 여기서, Zernike 모멘트는 회전에 무관한 특징을 갖는 변환으로 알려져 있다. 그러나 이 변환은 이동과 스케일 변화에 영향을 받는 특징을 가지고 있다. 따라서 본 논문에서는 전 두 단계에서 이동과 스케일에 무관한 특징을 갖는 정규화된 DFT 에너지 평면을 구했다. 이렇게 구해

진 평면에 회전에 무관한 것으로 알려진 Zernike 변환을 적용하여 이동, 스케일, 회전에 무관한 특징을 추출할 수 있다.

본 논문에서는 총 64개의 특징값을 가지고 있다. 추정된 cut-off(80%)값과 10차에 해당하는 Zernike 모멘트의 절대값 35과 26개에 해당하는 DFT의 에너지의 평균값 그리고 전체 DC와 AC에너지 특징들이 그것이다.

텍스처 특징 추출부에서는 주지한 바와 같이 입력 영상의 특징을 추출하게 되며, 데이터 베이스 구축부에서는 이를 데이터 베이스에 저장한다. 다음으로, 질의 영상이 입력될 경우 상기 텍스처 특징 추출부에서 특징들을 추출하게 되고, 거리 계산부에서는 데이터 베이스의 특징값들과 상기 추출한 질의 영상의 특징들을 비교하여 최소의 거리를 내는 데이터 베이스내의 영상을 찾고, 최소 거리 출력부는 그 최소 거리를 갖는 영상들을 출력함으로써, 사용자는 검색된 텍스처 영상을 신속하게 볼 수 있다.

다음으로, 사용자가 회전 및 스케일에 불변인지 아닌지에 대한 모드를 지정하여 텍스처 영상을 검색할 수 있는 방법도 가능하며, 이러한 사용자 모드 선택에 따른 텍스처 영상 검색 과정을 보여준다. 사용자 지정 모드는 네 개의 모드가 있으며, 이는 다음과 같다. 첫번째 모드는 회전 및 스케일에 불변인 텍스처를 검색할 수 있는 모드이며, 두 번째는 회전에는 관계없지만 스케일이 같은 텍스처를 검색하는 모드이다. 그리고 세번째 모드는 회전각은 같지만 스케일에 관계없는 텍스처를 검색할 수 있는 모드이며, 마지막으로 네번째는 회전각 및 스케일이 같은 텍스처를 검색할 수 있는 모드이다.

여기서 각각의 모드에 대한 알고리즘은 다음과 같다.

<회전 및 스케일에 불변인 검색법>

$D_{scale} = \min(f_{cut \cdot off, q}, f_{cut \cdot off, d}) / \max(f_{cut \cdot off, q}, f_{cut \cdot off, d})$
 If ($D_{scale} \geq 0.5 - d$)

$$Dis = \frac{\sum_{i=1}^{35} \|ZM_{i,q} - ZM_{i,d}\|}{\sqrt{std(ZM_i)}} + \frac{\|P_{0,q} - P_{0,d}\|}{\sqrt{std(P_0)}} + \frac{\|P_{AC,q} - P_{AC,d}\|}{\sqrt{std(P_{AC})}}$$

else Dis = Infinite

<회전에만 불변인 검색법>

If ($D_{scale} \geq 1 - d$)

$$Dis = \frac{\sum_{i=1}^{35} \|ZM_{i,q} - ZM_{i,d}\|}{\sqrt{std(ZM_i)}} + \frac{\|P_{0,q} - P_{0,d}\|}{\sqrt{std(P_0)}} + \frac{\|P_{AC,q} - P_{AC,d}\|}{\sqrt{std(P_{AC})}}$$

else Dis = Infinite

<스케일에만 불변인 검색법>

If ($D_{scale} \geq 0.5 - d$)

$$Dis = \frac{\sum_{i=1}^{35} \|ZM_{i,q} - ZM_{i,d}\|}{\sqrt{std(ZM_i)}} + \frac{\sum_{i=1}^{26} \|P_{i,q} - P_{i,d}\|}{\sqrt{std(P_i)}} + \frac{\|P_{0,q} - P_{0,d}\|}{\sqrt{std(P_0)}} + \frac{\|P_{AC,q} - P_{AC,d}\|}{\sqrt{std(P_{AC})}}$$

else Dis = Infinite

<회전 및 스케일이 같은 텍스처 검색법>

If ($D_{scale} \geq 1 - d$)

$$Dis = \frac{\sum_{i=1}^{35} \|ZM_{i,q} - ZM_{i,d}\|}{\sqrt{std(ZM_i)}} + \frac{\sum_{i=1}^{26} \|P_{i,q} - P_{i,d}\|}{\sqrt{std(P_i)}} + \frac{\|P_{0,q} - P_{0,d}\|}{\sqrt{std(P_0)}} + \frac{\|P_{AC,q} - P_{AC,d}\|}{\sqrt{std(P_{AC})}}$$

else Dis = Infinite

III. 실험결과

본 논문에서는 T1, T2, T3, T4, 그리고 T7의 5쌍의 실험영상을 사용하여 검색률을 통하여 제안한 알고리즘이 효과적임을 보였다. 여기서 T1은 116개의 영상으로 구성된 Brodatz 텍스처 영상을 사용하였으며, T2는 52개의 클래스로 구성된 ICU (Information Communication University)에서 제공된 영상이다. T3는 T1과 T2영상을 여러 각도로 회전시켜 얻은 영상으로 회전에 관한 실험을 위한 영상이다. T4는 스케일을 시험하기 위한 영상으로 스케일 변화를 점차로 증가시켜 검색률을 얻었다. 마지막으로 T7은 회전과 스케일이 동시에 된 영상으로 이 영상을 Corel의 텍스처 영상 300장에서 선택하여 만들어진 영상이다.

1) Brodatz (T1)

T1 데이터 베이스는 109개의 Brodatz 텍스처영상과 7 USC영상으로 구성되어 있다. 이들은 각각 512x512의 크기를 가지고 있다. 텍스처 검색률 계산을 위하여 본 실험에서는 각각의 영상을 16개의 16x16개의 영상으로 분할한 후 각 분할된 영상으로 나머지 영상을 찾는 개수를 측정한다.

2) ICU (T2)

이 데이터 베이스는 52개의 텍스처 영상으로 ICU에서 디지털 카메라로 취득한 영상이

다. 이 데이터 베이스도 Brodatz영상과 같은 방법으로 분할한 후 검색율을 계산하여 제안한 알고리즘의 성능을 알아본다.

3) 회전된 T1과 T2 (T3)

이 데이터 베이스는 T1에서 30개 T2에서 25개의 영상을 선정하여 T1의 경우 0, 10, 20, ... 90도로 회전시켜 영상을 만들고, T2의 경우는 0, 15, 30, 45, ..., 165도로 회전시켜 회전된 텍스처를 구성한다. 여기서 각각의 회전된 영상을 질의 영상으로 하여 다른 회전된 영상을 얼마나 찾는지를 검색율로 한다.

4) 스케일된 T1 (T4)

이 데이터 베이스는 T1 데이터 베이스를 스케일의 양에 따라 10개의 데이터 베이스를 구축한다. 다음과 같은 규칙에 따라 데이터 베이스가 구성된다.

$$\text{Rule1: } T4(i) = T4(i-1) \cup \text{Brodartz}(100-i*5)\%scaled \cup \text{Brodartz}(100+i*5)\%scale, \\ (i = 1, 2, \dots, 10)$$

$$\text{Rule2: } T4(0) = 115 \text{ original Brodatz}$$

여기서 Brodarz(x)%scaled는 Brodatz영상을 x%로 스케일한 영상 집합을 말한다. 결과적으로 Rule 1은 T4(i-1)영상 집합에 스케일된 영상의 집합을 합하는 방법으로 구성된다.

다음 표 1에서 제안한 알고리즘에 의한 실험결과를 보이고 있다. 표에서 보듯이 제안한 알고리즘이 이동, 스케일, 및 회전에 관계없이 좋은 결과를 낼 수 있다.

Database	특성	검색율
T1	Brodatz	71.4
T2	ICU	85.6
T3	Brodatz	100.0
	ICU	93.3
T4	a	97.8
	b	98.0
	c	97.8
	d	97.7
	e	97.5
	f	96.7
	g	95.4
	h	93.8
	i	91.8
	j	90.8
T7	Corel (70%)	99.2

표 1. 제안한 방법에 의한 텍스처 검색률

IV. 결론

본 논문에서 DFT와 Zernike 변환의 특성을 적절히 활용함으로써 이동, 스케일 및 회전에 무관한 텍스처 영상의 특징 추출이 가능하며, 이로써 이동, 스케일 및 회전에 관계없는 텍스처 영상의 검색이 가능함을 보였다. 또한, 기존의 많은 텍스처 영상 검색 방법들이 많은 필터를 사용하거나 통계적 모델을 적용하여 매우 많은 계산량 및 계산 시간을 요구하는데 반해, 제안한 방법은 DFT와 Zernike 변환만으로 텍스처 영상을 신속하게 검색할 수 있으며, 아울러 텍스처 영상 검색을 위한 비용도 절감할 수 있는 이점이 있다. 또한 변형에 강건한 텍스처 영상을 검색할 수 있으므로, 보다 인간의 시각적 영상 인식과 유사하여 멀티미디어 데이터 검색시 편리함을 도모해주는 이점도 있다.