FFT와 초점정보를 이용한 상대적 깊이지도의 생성

이진용*, 조진수*, 이일병*
*연세대학교 컴퓨터과학과
e-mail: *topjin55@hotmail.com

Generation of the Relative Depth Map using FFT and Focal Information

Jinyong Lee*, Jinsu Jo*, Yillbyung Lee*
*Dept. of Computer Science, Yonsei University

요 약

인간은 초점정보를 이용하여 단안만으로도 공간의 깊이를 지각할 수 있다. 이것은 한 번에 하나의 대상물에만 초점을 맞출 수 있고 그 외의 부분은 흐림 현상을 유도함으로써 이루어진다. 이는 초점이 맞는 대상물체로부터 멀어지면 멀어질수록 흐림 현상이 강해지는 원리를 이용한 것으로 주파수 성분의 변화량에 대한 연산과 깊은 관련이 있다. 본 논문에서는 이와 같은 인간의 시각 시스템의 요소 중 하나인 초점정보를 모방하여 초점거리가 다른 각각의 이미지들에 각각의 가중치를 부여하였다. 그리고 각 이미지들을 일정 블록으로 각각 분할하여 초점이 가장 잘 맞는 블록을 찾아내어 하나의 이미지로 통합하였다. 이때 각 영역은 자신이 속했던 이미지의 가중치를 따르게 한다. 각 이미지에서 가장 포커스 수치가 높은 영역을 찾기 위한 방법으로 주파수 영역 기반 처리와 공간 영역 기반 처리를 결합 하였다. 주파수 기반으로는 FFT(Fast Fourier Transform)에서 고주파 부분의 영역을 뽑아내어 포커스수치를 계산하였으며, 공간 영역 처리 기반으로는 이웃픽셀과의 차이가 임계값이하인 것을 제외한 영역을 뽑아내어 저주파 영역의 연산을 제거하는 방법과 단순히 Laplacian measure만을 사용하여 저주파까지도 포함한 방법의 두 가지를 적용하였다. 최종적으로 3개의 포커스 측정값을 결합시켜 포커스 수치를 계산한 후 각 블록의 가중치에 맞게 하나의 이미지로 통합하여 상대적 깊이지도를 생성하였다.

1. 서론

유비쿼터스(Ubiquitous)의 시대가 다가오면서 로봇과 컴퓨터 비젼(Computer Vision)에 대한 관심이 계속하여 커지고 있다. 로봇의 눈의 역할을 하는 컴퓨터 비젼(Computer Vision) 기술은 단순한 정보만을 획득할 수 있었던 센서(Sensor) 기술을 보완할 기술로 기대되며 이는 보다 많은 정보로 연산 처리가 가능한 시스템들이 등장하게 함으로써 인간생활에 도움을 주게 될 것이다. 이와 같은 컴퓨터 비젼(Computer Vision) 기술에서 사용되는 렌즈는 인간의 눈과 유사한 점이 많아 인간의 시각 시스템을 모방하는 것은 비젼(Vision) 기술의 발달에 큰 영향을 끼칠 것이라 기대된다. 그러나 인간의 시각시스템의 원리는 아직 완전히 밝혀지지 않았다. 따라서 어떤 방식으로 자연영상물을 인식하고 가공하는지는 정확히 알 수는 없다. 그러나 여러 가지 단서를 이용하여 인간의 시각시스템을 부분적으로 나마 유추해 볼 수는 있다. 그 중 대표적인 것으로 양안시차, 운동시차, 대기의 흐림 정도의 차이, 결의 기울기, 소실점, 채도의 차이, 초점정보 등이 있다. 양안시차 기법은 카메라가 두 대여야하며 가까이에 있는 물체에만 정확도가 높기 때문에 시스템을 최적화시키기에는 불리하다. 이에 반해 본 논문에서 다룰 카메라 한 대만을 사용하는 초점정보는 거리의 제약이나 비용 면에서 양안보다 유리하다. 초점정보를 이용한 깊이 측정 방법에는 초점이 잘 맞는 정도의 차이로 깊이를 지각하는 Depth from focus 방법[3,6,7]과 초점의 흐림 정도의 차이를 이용하는 Depth from defocus방법[1,2,3,4,5]이 있다. Depth from focus 방법은 한 번에 초점이 잘 맞는 하나의 대상물의 거리만을 알 수 있기 때문에 여러 장의 이미지로부터 전체적 거리정보를 계산해낸다.[3,6,7,] 이에 반해 Depth from Defocus는 초점이 잘 맞는 이미지와 초점이 맞지 않는 단지 몇 장 정도의 이미지로부터 흐림의 정도를 연산하여 거리정보를 뽑아낸다.[1,2,3,4,5] 본 논문에서는 초점거리가 다른 이미지들을 확보하여 각 이미지에 초점거리에 맞는 가중치를 적용한 후 이미지 각각 마다 일정 블록으로 분할하여 주파수 기반에서 FFT(Fast Fourier Transform)를 이용하고 공간 기반에서는 Laplacian

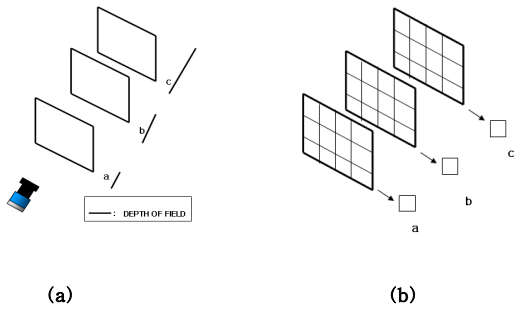
* 본 연구는 산업자원부의 뇌신경정보학사업의 지원을 받아 수행되었음

measure와 임계값을 고려한 Laplacian measure를 결합하여 포커스 수치를 계산하여 가장 높은 수치를 나타내는 블록을 해당 이미지의 가중치에 맞추어 하나의 이미지로 통합하여 상대적 깊이지도를 생성하는 알고리즘을 제안한다.

2. FFT(Fsat Fourier Transform)와 초점정보를 이용한 상대적 깊이지도 생성 알고리즘

2.1. 에지(Edge)추출 및 이미지에 가중치 부여

가까운 거리부터 먼 거리까지 초점을 바꾸어가면서 이미지를 획득한다. (그림 1-a)와 같이 피사계심도(Depth of field)를 고려하면서 획득된 각 이미지에 a, b, c의 가중치를 부여하는데 이때 가중치에 크기는 a>b>c의 관계가 성립하도록 한다. 따라서 가까운 곳에서 멀어질수록 가중치를 줄어들어 전체적으로 위에서부터 내려가는 계단의 형상을 가지도록 한다. 획득된 이미지는 모두 식(1)과 같은 DOG(Difference of Gaussians) 필터 거치 도록하여 잡음을 줄이면서 에지를 더욱 굵게 만든다. DOG는 라플라시안(Laplician)보다 잡음에 덜 민감하고 LOG(Laplician of Gaussian)보다 속도가 더 빠르며 σ^1, σ^2를 변화시킴으로써 에지(Edge) 검출 넓이를 조절할 수 있는 장점이 있다. DOG로 회선된 후 잡음이 줄어들고 에지(Edge)가 굵어진 여러 장의 이미지들을 각각 (그림1-b)과 같이 원하는 블록 크기로 잘라서 윈도우 단위로 처리하기 위한 준비를 한다. 이는 각 이미지에서 같은 위치에 있는 블록들끼리 포커스 수치를 계산하여 최고 수치의 가중치를 가지는 영역을 선택하여 하나의 이미지로 통합하기 위해서이다.



(그림 1) (a)초점거리가 다른 이미지의 획득, (b) 획득된 이미지들의 영역분할 및 분할된 영역의 가중치 적용

$$DoG(x,y) = \frac{e^{\frac{-(x^2+y^2)}{2\pi\sigma_1^2}}}{2\pi\sigma_1^2} - \frac{e^{\frac{-(x^2+y^2)}{2\pi\sigma_2^2}}}{2\pi\sigma_2^2} \quad (1)$$

2.2. FFT(Fsat Fourier Transform)를 이용한 포커스 수치 계산

x=0, 1, 2....., M-1, y=0, 1, 2,....., N-1인 $f(x,y)$가 M*N영상을 나타낼 경우 $F(u,v)$로 표기 하는 f의 2-D DFT(Discret Fourier Transform) 변환은 식 (2)와 같이 주어진다.

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux/M+vy/N)} \quad (2)$$

여기서 u =0, 1, 2, ..., M-1, v= 0, 1, 2, ..., N-1이다. 주파수 도메인은 단순히 u, v를 (주파수)변수로 하는 $F(u,v)$가 점유 하는 좌표 시스템이다. 이는 x, y를 (공간) 변수로 하는 $f(x,y)$가 점유하는 좌표 체계인 공간 도메인과 비슷하며 u, v에 의해 정의되는 M*N 직사각형 영역을 주파수 직사각형이라고 부르기도 한다. 이와 같은 DFT(Discret Fourier Transform)는 M*M의 크기를 가진 영상은 $(M^2)^2$의 복잡도를 가지기 때문에 FFT(Fsat Fourier Transform)를 사용하여 복잡도를 2M으로 줄인다.

FFT(Fast Fourier Transform)에 의하면 가운데에 고주파가 몰려있으며 외곽으로 갈수록 저주파가 강해진다는 점에 착안하여 그림(3)과 같이 각 이미지에서 획득된 같은 위치의 블록 마다 FFT(Fast Fourier Transform)로 바뀐 영역을 그림과 같이 9등분을 하여 고주파가 몰려 있는 가운데 부분만 추출한다. 그 뒤 Laplacian measure를 사용하여 포커스 수치를 계산하였다.

(그림 2) 각 이미지 블록의 FFT(Fast Fourier Transform) 적용하여 고주파 영역을 뽑아내는 과정

포커스 수치계산에 Laplician을 기본을 삼은 이유는 물체의 특성에 가장 영향을 적게 받는 계산법이기 때문이다.

식(3)은 Laplician 공식이다. 해당 블록에서 이를 적용하여 식(4)과 같이 x방향으로 얻어지는 $f(x,y)$값과 식(5)와 같이 y방향으로 진행되는 $f(x,y)$값을 각각 누적시켜 식(6)과 같이 그 두 방향의 누계의 최종 합으로 포커스 수치를 산출해 내었다.

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \quad (3)$$

$$frequency_x = \sum_{1}^{rowcol-1}\sum_{2} f(x+1,y) + f(x-1,y) - 2f(x,y) \quad (4)$$

$$frequency_y = \sum_{1}^{rowcol-1}\sum_{2} f(x,y+1) + f(x,y-1) - 2f(x,y) \quad (5)$$

$$frequency_{fft} = frequency_x + frequecny_y \quad (6)$$

포커스 수치가 최대가 되는 위치가 물체와 렌즈간의 거리라고 추정할 수 있다.

2.3. 공간 영역을 기반으로 한 포커스 수치 계산

포커스 수치를 Laplacian measure를 사용 하여 계산을 할 때 그 결과 값이 임계값이하일 경우 저주파 영역이라 가정을 하고 제외 한다. 임계값 이상일 경우는 초점이 잘 맞는 고주파 영역으로 포커스 수치 계산에 포함 시킨다. 임계값 이상인 경우가 발생할 때 마다 카운트를 증가시켜 연산 뒤에 카운트로 고주파부분 연산이 얼마나 진행되었는지 정규화 시킬 수 있도록 한다.

다음 식(7) 은 x방향에 의한 연산을 나타낸다.

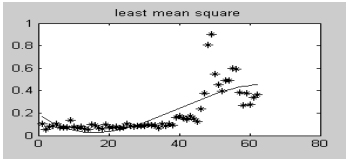
$$\begin{aligned} &\text{if} \\ &middle = f(x+1,y) + f(x-1,y) - 2f(x,y) \\ &then \\ &\ frequency_x = frequency_x + middle \\ &count_x = count_x + 1 \end{aligned} \tag{7}$$

y 방향도 식(7) 같은 방식으로 접근하고 최종적으로 식(8)과 같이 x방향의 frequency는 x방향의 count로 y방향의 frequency는 y방향의 count로 나누어 주고 두 개 의 값을 합쳐 포커스 수치를 계산한다.

$$\begin{aligned} &frequency_{area} = \\ &frequency_x / count_x + frequency_y / count_y \end{aligned} \tag{8}$$

2.4. 최적 블록 도출 및 이미지 정합

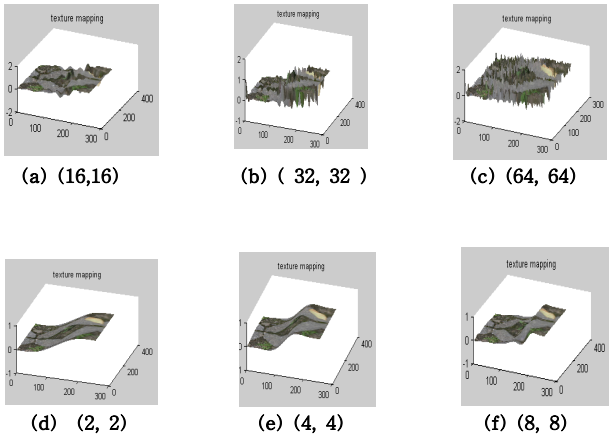
FFT(Fast Fourier Transform)로 계산된 포커스 수치와 임계값을 바탕으로 계산된 포커스 수치 그리고 임계값이 적용되지 않고 Laplacian measure만을 적용한 포커스 수치 3개를 결합시킨 뒤 최소 제곱 오차법으로 여러 이미지에서 같은 위치에 있는 블록들 중 가장 포커스가 잘 맞는 블록을 뽑아낸다. 뽑아져 나온 블록은 원래 자신의 이미지에 있던 초기 가중치 값을 바탕으로 하나의 이미지에 통합된다. 하나로 통합된 이미지의 블록들의 각 값을 0~1 사이로 정규화 시켜 상대적 깊이지도를 만들어 낸다. (그림 4)는 최소 제곱 오차법을 각 이미지 영역에 적용하여 최적의 블록을 뽑아내는 과정을 보여준다.



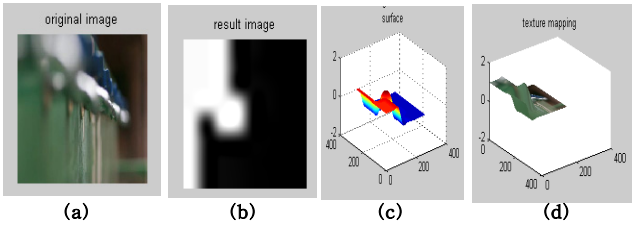
(그림 3) x축: 이미지 번호, y 축: 각 이미지에 상응하는 영역의 최종 포커스 수치 값, 각 수치 값의 최소 제곱 오차법 적용 과정

3. 실험 및 결과

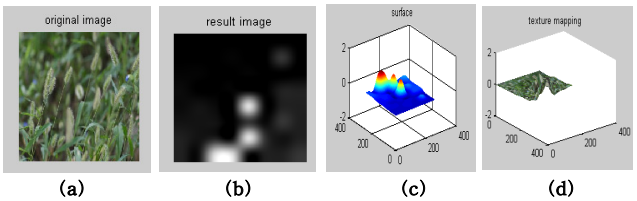
본 논문은 Canon EOS 400D와 F1.2렌즈를 이용하여 같은 공간에서 0.95m~ 무한대까지 거리에 초점을 조절하며 맞추어 이미지를 획득하였다. 이미지 크기는 256*256에서 시행되었으며, 실험은 62장, 8장, 4장의 이미지의 통합으로 상대적 깊이지도를 생성하도록 시행하였다. 카메라 초점을 수동으로 맞추었기 때문에 피사계심도(Depth of field)로 인한 에러를 줄이기 위해 최대 62장까지 잡아 보았다. 이미지 영역 분할은 (그림 4)의 결과처럼 블록의 크기가 너무 커질 경우 상대적 지도가 부드럽지만 세부 표현 되지 않았으면 너무 작아질 경우는 잡음이 강한 문제가 있어 전체 이미지를 8*8 영역으로 나누어 잡아 주었다.



(그림 4) (a) 16, 16의 블록 분할결과 (b)32, 32의 블록 분할결과, (c) 64, 64의 블록 분할결과 (d) 2, 2의 블록 영역 결과 (e) 4, 4의 블록 분할결과 (f) (8,8)의 블록 분할 결과

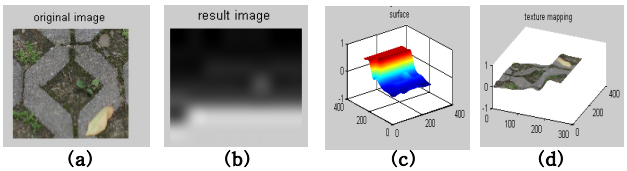


(그림 5) 실험 1. 62장 이미지의 통합, (a) 원본 이미지 모양, (b) 2차원으로 표현된 상대적 깊이지도, (c) 3차원으로 표현된 상대적 깊이지도 (d) texture mapping의 3차원의 상대적 깊이지도



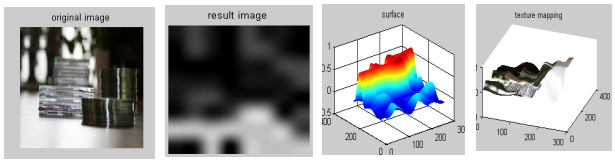
(그림 6) 실험 2. 8장 이미지의 통합, (a) 원본 이미지 모양, (b) 2

차원으로 표현된 상대적 깊이지도, (c) 3차원으로 표현된 상대적 깊이지도 (d) texture mapping의 3차원의 상대적 깊이지도



(a) (b) (c) (d)

(그림 7) 실험 3. 8장 이미지의 통합, (a) 원본 이미지 모양, (b) 2차원으로 표현된 상대적 깊이지도, (c) 3차원으로 표현된 상대적 깊이지도 (d) texture mapping의 3차원의 상대적 깊이지도



(a) (b) (c) (d)

(그림 8) 실험 4. 4장 이미지의 통합, (a) 원본 이미지 모양, (b) 2차원으로 표현된 상대적 깊이지도, (c) 3차원으로 표현된 상대적 깊이지도 (d) texture mapping의 3차원의 상대적 깊이지도

(그림 5)의 실험 1의 경우는 0.95m부터 무한대까지 초점을 변경시키면서 62개 범위의 이미지를 획득하였다. 이것은 수동 초점 조절에 따른 피사계 심도(Depth of field)의 오차를 최대한으로 줄이기 위해 62단계로 획득하여 복원한 결과이다. 0.95m에 가장 밝게 나타나면 주변으로 갈수록 가중치가 줄어들어 점점 어두워지며 이를 3차원 그래프와 Texture mapping으로 표현하였다.

(그림 6)의 실험2는 풀숲을 기준으로 8개의 범위로 잡아주었으며 (그림 7)의 실험 3도 위의 경우와 같이 나타내었다 (그림 8)의 실험 4는 4개의 범위로 피사계 심도(Depth of field)의 확연한 차이를 주어 실험을 하였다. 여기서는 물체는 1m, 1m 20cm 1m 50cm, 2m에 각 물체에 초점을 맞추고 상대적 깊이를 도출해 보았다. 실험 영상에 따라 결과는 달랐고 이는 textureless와 빛의 반사율 등의 잡음으로 여겨지며 전체적인 상대적 지도형성은 층을 이루면서 만족할 만한 수준으로 이루어졌다.

4. 결론 및 향 후 연구방향

컴퓨터 비젼(Computer Vision)은 인간의 시각 시스템과 매우 유사한 렌즈 시스템을 사용하고 있기 때문에 인간의 눈을 모방한 컴퓨터 비젼 시스템(Computer Vision System)을 구현하는 것은 다가올 로봇과 유비쿼터스(Ubiquitous) 세상의 발전에 크게 이바지 할 것으로 기대된다. 이에 인간의 시각 시스템 중 단안으로 깊이를 지각할 수 있는 초점 정보를 모방하여 상대적 깊이지도 시스템을 구현하였다. 이 시스템은 인간은 한 번에 하나의 대상에만 초점을 맞출 수 있다는 것에 착안하여 자연 영상에서 같은 공간의 이미지를 초점만 바꾸어 여러 장을 획득하여 각 초점에 따른 가중치를 부여하였다. 그리고 각 이미지를 일정 크기의 블록으로 분할하여 각 이미지에서 같은 위치에 속한 블록의 포커스 수치를 비교하여 여러 이미지 중 최적의 포커스 수치를 가진 블록을 찾아내어 하나의 이미지로 통합하였다. 최적의 포커스 수치를 계산하기 위하여 공간 영역에서는 임계값을 이용한 Laplacian measure와 단순히 Laplacian measure만 적용한 두 가지 포커스 수치 값을 산출하였으며 주파수 영역처리에서는 FFT(Fast Fourier Transform)를 이용한 고주파 영역에서의 Laplacian measure를 적용한 포커스 수치 값을 산출하였다. 그 뒤 3가지 포커스 수치 값을 결합하여 여러 이미지에서 최소 제곱 오차법을 사용하여 최적의 포커스 수치를 가지는 블록들만을 뽑아내어 하나의 이미지로 통합을 하였다. 이때 각 블록은 자기가 속한 이미지의 가중치 값을 따르게 됨으로 깊이를 표현할 수 있다. 이를 3차원으로 표현하여 상대적 깊이지도를 완성하였다. 실험 결과는 대체적으로 만족할 만한 수준으로 나타났다. textureless와 빛의 반사에 따른 잡음 부분을 해결한다면 더욱 좋은 결과를 가질 수 있을 것이라 예상된다. 따라서 보다 많은 경우를 가정한 실험과 보완이 진행해야 할 것이다. 또한 초점을 이용한 정보가 다른 시각적 요소와의 통합을 이룬다면 보다 많은 향상을 일으킬 것으로 기대된다.

참고문헌

[1] A. Pentland, "A new sense of depth of field", Proc. IEEE Trans. PAMI, pp.523-531, 1987

[2] Subbarao. M., Gurumoorthy. N. "Depth recovery from blurred edges", Proc. Computer Vision and Pattern Recognition, pp.498-503, June 1998

[3] Y. Xing., S. A shafer, "Depth from focusing and defocusing", CVPR, pp.68-73, 1993

[4] Subbarao. M., Wei. T.-C., Surya,.G., "Focused image recovery from two defocused images recorded with different camera settings" Image Processing, IEEE Transactions on Vol(4), Issue 12, pp.1613 - 1628, Dec. 1995

[5] Schneider. G., Heit. B., Honig. J., Bremont. J., "Monocular depth perception by evaluation of the blur in defocused images", Image Processing, IEEE International Conference, Vol(2), pp.116 - 119 , Nov. 1994

[6] Subbarao, M, "Parallel Depth Recovery By Changing Camera Parameters", Computer Vision., 1988. Second International Conference, pp.149-155, Dec. 1988

[7] Nourbakhsh, D. Andre, C. Tomasi, and M. Genesereth, "Mobile Robot Obstacle Avoidance via Depth from Focus," Robotics and Autonomous Systems, Vol(22), pp. 151-158, June, 1997