

단일시점 영상의 소실점을 기반한 깊이

정보 예측과 영상 합성 방법

백은뜸, 호요성

광주과학기술원 정보통신공학부

eutteum@gist.ac.kr , hoyo@gist.ac.kr

초록: 단일시점 영상으로 깊이 값을 예측하는 것은 어렵고 까다로운 작업이다. 특히, 단일시점 영상에서 소실점을 찾는 일은 어려운 일이며, 촬영 환경에 따라 결과에 많은 영향을 준다. 이 논문은 단일시점 외부 영상으로 깊이 값을 예측하고, 예측된 깊이 값을 이용하여 영상을 합성하는 방법을 제안한다. 영상을 이진화하여 경계선과 배경으로 나누고, 잡음을 제거하여 소실선과 연관된 강인한 선을 찾는다. 소실선을 이용하여 소실점을 구한 뒤, 하늘과 지면으로 분리된 영상과 소실점을 활용하여 깊이 영상을 만든다. 마지막으로 깊이 영상을 활용하여 영상을 합성한다. 실험을 통해 거의 정확한 소실점을 예측했고, 깊이 영상을 활용한 합성 영상도 좋은 결과를 얻었다.

주제어: Vanishing point; Depth estimation; Hough transform; Canny edge detection; Image synthesis.

I. 서론

스테레오 영상을 사용하여 깊이 정보를 예측하는 방법은 널리 사용되고 있으며, 정확도 역시 매우 높다. 하지만 단일시점 영상으로 깊이 정보를 찾아내는 것은 어려운 일이고, 많은 제약사항이 존재한다. 때문에 소실점을 활용하여 깊이 값의 정보를 찾는 연구가 진행되고 있다. 일반적으로 거리에 상관없이 그 형태가 유지되는 점과 선의 성질과 직선의 교차점의 정보를 사용하여 소실점(vanishing point)을 예측한다. 소실점이란 3D공간에 있는 평행한 선들의 집합이 핀 홀 카메라로 얻어진 2D공간에 투영될 때 영상 면에서 수렴하고 만나는 점으로 영상에서 가장 먼 부분을 의미한다. 이전의 소실점에 대한 연구는 대부분 영상에서 분할된 선을 활용하여 소실점을 찾았다. T. Tuytelaars는 화소의 변화도를 사용하여 소실점을 찾는 연구를 했고 [1], Barnard는 카메라의 시각적 중심에 있는 기우시안 구체를 제안했고 [2], Van den Heuvel은 영상에서 상호적으로 직교하는 방향에 대해 소개하였다 [3]. 하지만, 소실점에 관련된 직선이 없거나, 주변에 직선을 찾기 어려운 환경에서 소실점을 예측하기는 어려운 일이다.

단일시점 영상은 깊이 정보를 갖고 있지 않기 때문에 깊이 정보를 이용한 3차원 변형이나, 합성 등을 할 때 어려움이 있다. 그래서 단일시점 영상에 존재하는 깊이 단서를 이용하여 깊이 정보를 예측하는 방법에 대한 연구가 진행되고 있다. Battiato는 색상 정보와

깊이 영상의 관계를 활용하여 깊이를 예측하였고 [4], S.A. Valencia은 포커스에 따라 달라지는 영상의 특성을 활용하여 깊이를 예측하였다[5]. 기존에 연구들에서는 객체를 구분하지 않고 소실점을 기반으로 하여 정해진 패턴의 깊이 정보를 생성하거나, 객체를 종류별로 구분하지 않고 하나의 깊이 값을 객체에 일괄적으로 할당하는 연구가 진행되었다. 이러한 방법은 객체의 수가 많아지면 깊이 값이 정교하지 못하다는 단점을 갖고 있다. 특히 객체가 먼 곳에서부터 가까운 곳까지 연속하여 존재하는 경우에 정확성은 더욱 감소하게 된다.

본 논문은 단일시점 외부 영상을 사용하여 정확한 소실선을 검출하여 소실점을 예측하고, 영상분할 방법을 사용하여 하늘과 지면을 분리하여, 소실점 정보와 분할된 영상을 활용하여 알맞은 깊이 정보를 예측한다. 또한 예측된 깊이 정보를 활용한 정교한 합성 방법을 소개한다.

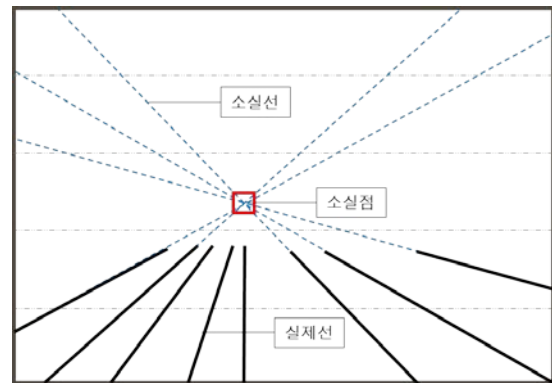


그림 1. 소실점과 소실선의 관계

II. 소실점 예측

1. 소실선 검출을 위한 전처리

단일시점 영상의 원근 정보는 깊이 값을 예측하는데 중요한 정보이다. 그림 1에서 보는 것과 같이 영상 내의 직선들을 무한하게 연장하여 교차되는 점을 소실점이라고 하고, 소실점과 만나는 직선들을 소실선이라고 한다. 하지만 영상에서 모든 직선이 소실점에 교차되지 않는다. 그리고 소실점에 관련된 직선이라고 할지라도 실제 영상에서 모든 소실선이 소실점에 교차하지

않는다. 때문에 정확한 소실선을 검출하고, 소실점과 관련이 없는 직선이나 잡음을 제거하는 것은 중요하다.

본 논문에서는 에지 성분을 검출하기 위하여 캐니 에지 추출 방법을 사용하였다[6]. 추출된 에지 영상은 모폴로지 기법인 침식과 팽창을 사용하여 직선과 관련 없는 잡음을 제거하였다. 침식과 팽창연산은 순서가 중요하다. 순서를 지키지 않으면 직선 정보를 잃어 버릴 수 있다.

2. 직선 검출

잡음이 제거된 에지 영상에 허프 변환을 사용하여 직선을 검출한다. 허프 변환은 한 평면 위에 놓여 있는 점들의 집합 $P(x, y)$ 를 지나는 모든 직선에 대한 식이다. 식 (1)에서 보듯 허프 변환은 극좌표계 형태의 직선의 방정식을 사용한다.

$$r = x \cos \theta + y \sin \theta \quad (1)$$

그림 2에서 r 은 직선과 영상 원점간의 거리이고, θ 는 선에 대한 수직 각도임을 보여준다. 허프 변환은 기존에 상수로 사용되던 θ 와 r 을 변수로 두고, 변수로 사용되던 x 와 y 를 상수로 사용하여, r, θ 평면에서의 곡선의 교차점을 얻어 직선을 얻는다. 하지만 허프 변환은 단순히 영상에서 정렬된 에지 화소를 찾기 때문에 일부 잘못된 감지를 하거나 소실선에 관련이 없는 직선이 검출되기도 한다. 따라서 영상의 크기에 따라 선의 길이와 간격 그리고 누적된 점의 최소 개수를 설정하여 정확한 직선을 검출한다. 누적된 점의 개수란 x, y 의 값이 일정 할 때, r, θ 를 변수로 두어 생기는 곡선의 누적된 값을 뜻한다.

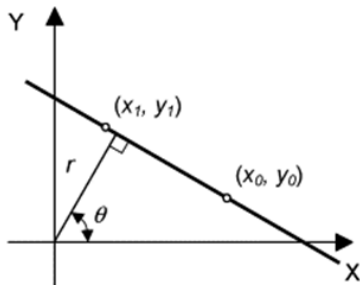


그림 2. 극 좌표계

$$\begin{cases} r_0 = x \cos \theta_0 + y \sin \theta_0 \\ \vdots \\ r_{n-1} = x \cos \theta_{n-1} + y \sin \theta_{n-1} \\ r_n = x \cos \theta_n + y \sin \theta_n \end{cases} \quad (2)$$

3. 소실점 예측

허프 변환으로 얻어진 직선을 사용하여 소실점을 예측한다. 얻어진 직선들은 식 (2)에서 보듯 θ 와 r 을 가지고 있는 직선의 방정식으로 표현 할 수 있다. 먼저 두 개의 방정식의 관계에서 해를 찾아 누적한다. 단, 각각의 방정식은 해가 무수히 많거나, 해가 없거나, 하나

의 해를 갖게 된다. 때문에 해가 무수히 많거나, 해가 없는 경우는 누적에 포함시키지 않는다.

두 방정식의 해를 얻기 위하여 그림 3의 알고리즘 순서로 해를 찾아 값을 배열 $P[x][y]$ 에 누적한다. 누적된 값의 분포를 사용하여 소실점을 예측한다. 하지만 영상에서 모든 소실선이 한 점에 교차되지 않는다. 때문에 누적된 값이 큰 부분과 교차점이 많이 발생하는 부분이 소실점이 될 것이라고 가정하고 소실점을 예측한다. 우리는 교차점이 많이 발생하는 부분을 찾기 위해서 누적된 점 x, y 에서 나머지 다른 점까지의 유클리드 거리(Euclidean distance)를 각각 계산하여 합이 가장 작은 부분을 교차점이 많이 발생하는 부분이라고 가정 할 것이다. 식 (3)의 $W(\cdot)$ 는 유클리드 거리를 나타내는 식이다. 식 (4)는 한 점에서 다른 점까지의 유클리드 거리를 각각 계산하여 합하는 식이다.

$$W(P_0, P_k) = \sqrt{(x_0 - x_k)^2 + (y_0 - y_k)^2} \quad (3)$$

$$\sum_{k=c}^n W(P(x_p, y_p), P(x_k, y_k)) \quad (4)$$

n 은 교차된 점의 개수이고, p 는 교차점, 그리고 q 는 p 에 속하지 않은 나머지 교차점을 의미한다.

```

P[x][y]; // set the entire array to 0
n // the number of equations

for i=0 to n-1;
  for j=i+1 to n;
    if ( an infinite number of solutions)
      then break;
    if ( no solution)
      then break;
    if (one solution)
      then (find the solution (x, y) for the i-th and j-th
            equations)
            P[x][y]++;
    
```

그림 3. 방정식 해 누적 알고리즘

III. 깊이 정보 예측과 합성 방법

1. 깊이 정보 예측

실외 영상에서 사진을 촬영할 때, 일반적으로 영상의 상단은 하늘을 포함하고, 하단 부분은 지면을 포함한다. 그래서 하늘은 카메라로부터 거리를 무한히 멀다고 가정 하고, 영상의 하단부분은 카메라와 가장 가깝다고 가정 한다. 이러한 가정과 예측된 소실점 정보를 가지고 배경의 깊이 정보를 예측한다. 깊이 정보를 예측하기 위해 색상 영상을 흑백 영상으로 바꾸어 k-means segmentation 방법으로 영상을 분할한다[7].

그림 4는 하늘부분과 지면부분이 분리된 영상을 보여준다. 모폴로지 기법을 활용하여 후 처리를 한다. 하늘을 포함하는 부분은 흑백 영상에서 0을 삽입하고, 지면 부분은 세 부분으로 나누어 깊이 값을 할당한다. 나누어진 세 부분에 대해 각각 기준점으로부터 순차적으로 값을 삽입한다. 먼저, 지면 2 부분은 식 (5)를 사용하여 깊이 값을 넣는다. y_{max} 는 영상의 가장 밑 y 값이고, y_{van} 은 소실점의 y 값이고, y_n 은 현재 좌표의 y 값이다. 지면 1과 지면 3은 영상 하단의 꼭지점에서 소실점까지 이어진 직선 위의 지면 2 부분의 깊이 값과 하늘과의 경계선까지 순차적으로 깊이 값을 할당한다.

$$I(x, y_n) = \frac{255}{y_{max} - y_{van}} * (y_n - y_{van}) \quad (5)$$

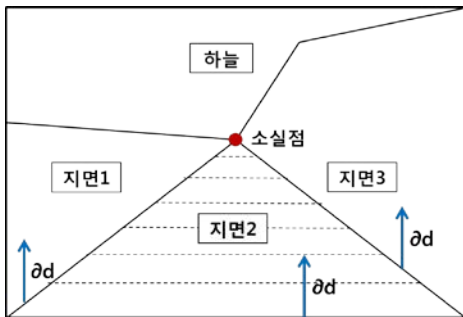


그림 4. 깊이 값 할당을 위한 영상분할

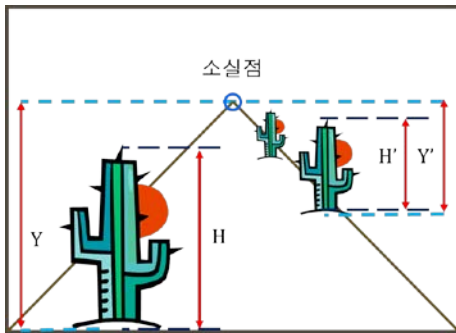


그림 5. 거리와 객체의 높이에 따른 비율

2. 예측된 깊이 정보를 활용한 영상 합성

일반적으로 사진을 합성할 때, 객체의 크기와 원근감을 고려하지 않고 영상을 합성하게 된다. 하지만 원근감이 고려되지 않은 영상은 부 자연스러운 합성 결과를 보여준다. 때문에 주어진 전경 영상의 크기와 거리의 비율을 고려해야 한다. 합성할 전경의 크기는 가장 가까운 위치에 있을 때의 크기라고 가정한다. 거리와 높이의 비율로 전경의 크기를 조절한다. 그림 5는 거리와 객체의 높이에 따른 비율을 보여준다. Y는 소실점에서 영상의 끝부분까지의 거리, H는 입력한

영상의 크기, Y'는 합성할 위치 그리고 H'는 합성될 크기를 의미한다. $H : Y = H' : Y'$ 의 비례식을 가지며 H'를 구하여 영상을 합성한다.

IV. 결과

실험은 480*352 크기의 영상을 사용하였다. 그림 6은 분할된 영상의 결과를 보여준다. k-means segmentation 방법으로 영상을 분할할 때 k는 2로 두어서 이진화된 영상을 만들어 사용하였다. 그림 7(a)는 원본영상이고 그림 7(b)는 그림 7(a)를 이용하여 직선을 검출한 영상이다. 직선을 검출하기 위해 에지성분으로 세그멘테이션을 수행하고, 허프 변환으로 직선을 검출하였다. 검출 시에 강인한 직선을 검출하기 위하여, 선분의 길이가 최소 40 화소 이상일 때의 직선을 강인한 직선이라고 가정하고 검출하였다. 하지만 영상의 크기에 따라 최소 선분 길이는 조정이 필요하다. 그림 7(c)는 검출된 직선으로 소실점을 예측한 영상이다. 아래의 영상은 (189, 167)이 소실점으로 예측되었다.

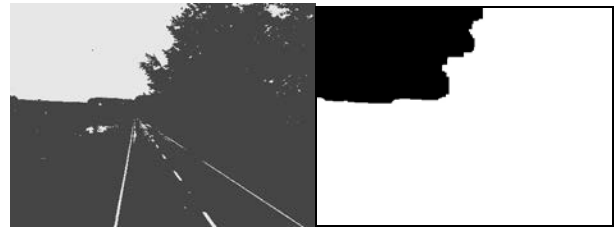
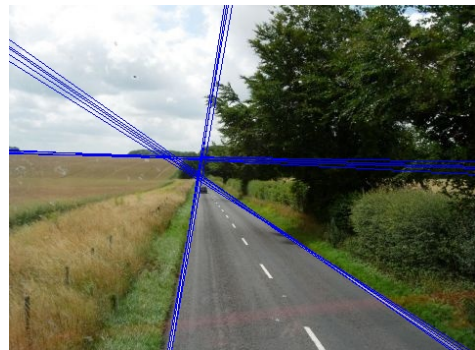


그림 6. K-means 영상분할



(a) 원본 영상



(b) 강인한 직선 검출



(c) 소실점 예측

그림 7. 소실점 예측

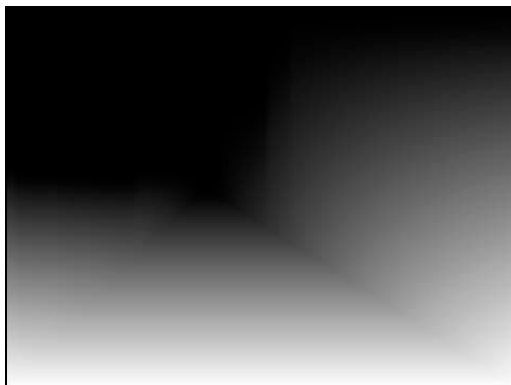


그림 8. 예측된 깊이 영상



그림 9. 예측된 깊이 정보를 활용한 영상합성

그림 8은 예측된 소실점과 이진 분할된 영상을 사용하여 깊이를 예측하여 얻은 깊이 영상을 보여준다. 그림 9은 예측된 깊이 정보를 활용한 영상합성을 보여준다. 합성하려는 객체의 원래 크기와 소실점으로부터 거리의 비율로 합성하여, 위치에 따라 객체의 크기가 변하는 것을 볼 수 있었다.

V. 결론

깊이 카메라나 스테레오 카메라를 이용할 경우 정

확한 깊이 영상을 얻을 수 있지만, 일반적으로 단일시점 색상 카메라로 촬영된 영상이 많기 때문에 단일시점 영상에서 깊이를 예측하는 연구는 중요하다. 본 논문에서는 단일시점 영상의 소실점기반 깊이 정보 예측과 합성 방법을 제안하였다. 소실점에 관련된 강인한 직선을 검출하고 소실점을 예측하고, 실외 영상의 특성을 고려하여 깊이 영상을 획득한다. 그리고 깊이 정보를 이용하여 객체를 합성하는 방법에 소개하였다. 실험결과를 통해 제안한 방법이 거의 정확한 소실점을 예측한 것을 알 수 있다. 하지만, 제안한 깊이 영상 예측방법은 영상 전체적인 깊이는 예측하였지만 객체의 정확한 깊이 값을 찾지 못함을 보여준다. 따라서, 객체에 대한 깊이 값은 연구가 더 필요하다.

감사의 글

이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2013-067321).

참고문헌

- [1] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. *The cascaded Hough transform as an aid in aerial image interpretation*. Proc. ICCV, pp. 67–72, Jan. 1998.
- [2] S. T. Barnard. *Interpreting perspective images*. Artificial Intelligence 21, pp. 435-462, 1983.
- [3] F. A. van den Heuvel: *Vanishing point detection for architectural photogrammetry*. International Archives of Photogrammetry and Remote Sensing, vol. XXXII part 5, pp. 652-659, 1998.
- [4] S. Battiato, S. Curti, E. Scordato, M. Tortora, and M. La Cascia, *Depth map generation by image classification*, Three-Dimensional Image Capture and Applications VI, vol. 5302, pp. 95-104, 2004.
- [5] S. A. Valencia and R. M. Rodriguez-Dagnino, *Synthesizing stereo 3D views from focus cues in monoscopic 2D images*, Proc of the SPIE, vol. 5006, pp. 377-388, 2003.
- [6] J. Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, issue 6, pp. 679-698, Nov. 1896.
- [7] C. W. Chen, J. B. Luo, and K. J. Parker, *Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications*, IEEE Trans. on Image Processing, vol. 7, pp. 1673–1683, Dec. 1998.
- [8] B. K. P. Horn. *Robot vision*, pp. 66-69, 299-333. MIT Press, 1986.