

로스 텍스처 필터 기반 영상의 자동 깊이 생성 기법

조철용 김제동 장성은 최창열 김만배
강원대학교 IT대학 컴퓨터정보통신공학과
{cheolyong, jedongkim, jse4485, changyeol, manbae}@kangwon.ac.kr

Automatic Depth Generation Using Laws' Texture Filter

Cheolyong Jo, Jedong Kim, Sung Eun Jang, Changyeol Choi and Manbae Kim
Dept. of Computer and Communications Engineering, IT College, Kangwon National University

요약

영상의 깊이 정보를 추출하는 것은 매우 어려운 연구이다. 다양한 유형의 영상 구조의 분석이 필요하지만 많은 경우에 주관적인 판단의 도움이 필요하다. 본 논문에서는 로스 텍스처 필터를 기반으로 정지 영상의 깊이를 자동으로 생성하는 방법을 제안한다. 로스 텍스처 필터는 단안 비전에서 3D 깊이를 얻기 위한 방법으로 활용되었는데, 실제 2D 영상에서 깊이를 예측하기 위해 텍스처 편차, 텍스처 기울기, 색상 등을 활용한다. 로스 필터는 1x5 벡터로부터 콘볼루션을 이용하여, 20여개의 5x5 콘볼루션 필터가 구해지는데, 영상에 필터를 적용하여 로스 에너지를 계산한다. 구해진 에너지를 깊이 맵으로 변환하고, 깊이 맵에서 특징 점을 구하고, 특징 점들로부터 텔러노이 삼각화를 이용하여 삼각형 깊이 메쉬를 얻는다. 구해진 깊이 맵의 성능을 측정하기 위해 카메라 시점을 변경하면서 영상의 3D 구조를 분석하였으며, 입체영상을 생성하여 3D 입체 시청 결과를 분석하였다. 실험에서는 로스 텍스처 필터를 이용하는 깊이 생성 방법이 좋은 효과를 얻는 것을 확인하였다.

1. 서론

컴퓨터 그래픽과 달리 실사 영상의 깊이를 예측하는 것은 어려운 작업이다. 스테레오 영상은 스테레오 정합과정을 통해 깊이를 생성할 수 있다 [1]. 그러나, 정지 영상의 깊이를 생성하는 방법은 영상 분석의 어려움으로 정확한 깊이 예측이 어렵다. Depth from shading, Depth from motion, Depth from edge 등의 다양한 Depth from X 기법이 있지만, 복잡도 및 제한된 환경에서만 가능한 방법들이다 [2~5]. 또한 이 기법들은 결과로 얻은 깊이 데이터에만 관심이 있고, 실제 입체영상을 생성하여 관찰하지 않아서, 실제 구해진 깊이가 실제 입체영상과의 관계를 연구하지 않았다. 따라서 본 논문은 로스 텍스처 필터(Laws' Texture Filter)를 기반으로 하는 자동 깊이 맵 생성 기법을 제안한다. 제안 방법은 2D 영상 내에 존재하는 전경 객체와 배경을 대상으로 추출된 깊이의 특성을 분석하고, 전경 객체와 배경을 분할할 수 있는 가능성을 조사한다. 구해진 깊이 맵을 이용하여 입체 영상을 생성하고 시청 결과를 분석한다.

본 논문의 구성은 다음과 같다. 2절에서는 로스 텍스처 필터를 소개한다. 3절에서는 로스 텍스처 필터를 이용하여 깊이 맵을 생성하는 방법을 제안한다, 4절에서는 실험의 결과를 보여주고, 결과 분석을 통해 제안 방법의 성능을 검증한다, 마지막으로 5절에서는 결론을 맺는다.

2. 로스 텍스처 필터

텍스처란 어떤 물질 자체의 특유의 재질을 나타내는 촉각적, 시각적 느낌을 말한다. 영상에서 텍스처란 화상 중에 존재하는 일정한 패턴이나, 서로 관련된 픽셀들이 그룹을 이루고 있는 형태 등을 말한다. 3차원 그래픽에서는 물체의 질감을 주기 위한 목재, 천, 가죽, 물의 표면 모양이나 그 물체의 도안 모양을 말한다.

인간은 3D 깊이감을 얻기 위해 단안 및 스테레오 정보로 구성된 다양한 종류의 시각 정보를 이용한다 [6]. 단안 정보 중 에지(edge) 방향의 분포를 알려주는 텍스처 그라디언트(texture gradient)는 깊이 인식의 중요한 역할을 담당한다. 로스 텍스처 필터는 단안 비전에서 3D 깊이를 얻기 위한 방법으로 종종 사용된다 [7][8]. 실제 2D 영상에서 깊이를 예측하기 위해 일반적으로 텍스처 편차(deviation), 텍스처 그라디언트(gradient), 색상 등을 활용하는데, 이의 구현을 위해서 로스 텍스처 필터가 활용되어 왔다 [9].

그림 1은 제안하는 로스 필터 깊이의 블록도를 보여준다. 먼저 영상을 휘도 성분인 Y 채널과 색차 성분인 Cb, Cr 채널로 분할한다. 조명의 영향을 감소하기 위해서 Y 채널 픽셀 주변의 평균값을 구한 후, 픽셀로부터 평균값을 차감한다. 다음으로 로스 콘볼루션(convolution) 필터를 적용하여 텍스처 에너지를 계산하고, 얻어진 에너지로부터 최종 깊이 맵으로 변환한다.

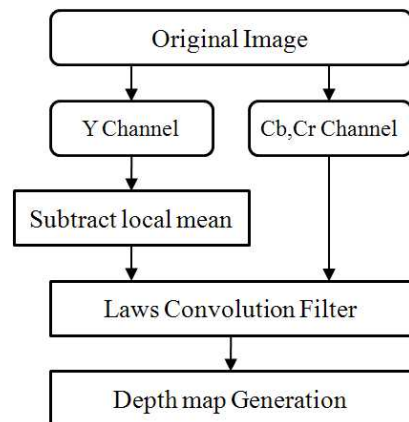


그림 1. 로스 필터 깊이의 블록도

2.1 기본 이론

2D 영상의 텍스처 에너지 측정은 콘볼루션 필터를 적용해서 얻어진다. 로스 텍스처 에너지는 텍스처의 평균 명암, 에지, 스포트(spot), 리플(ripple) 및 웨이브(wave)의 특성을 가지고 있다. 이 특성을 검출하기 위해 다음 3가지의 1x3 필터가 이용된다.

L3 = (1, 2, 1) (local average, level)

E3 = (-1, 0, 1) (first difference, edge)

S3 = (-1, 2, -1) (second difference, spot)

각 필터는 평균 검출 (local average or level), 에지 검출(edge detection), 스포트 검출(spot detection)을 하기 위한 계수이다. 필터의 확장을 위해 각 필터에 자신 및 다른 필터를 콘볼루션하면, 다음 5개의 1x5 필터가 얻어진다.

L5 = (1, 4, 6, 4, 1) (center weighted local average)

E5 = (-1, -2, 0, 2, 1) (row or column step edges)

S5 = (-1, 0, 2, 0, -1) (detects spots)

R5 = (1, -4, 6, -4, 1) (detects ripples)

W5 = (-1, 2, 0, -2, 1) (detects waves)

각 필터는 Local average(or Level), Edge, Spot, Ripple, Wave를 검출한다. L5를 제외한 모든 필터는 계수의 합이 0인 zero-sum 필터이다. 2D 콘볼루션 마스크 M은 상기 1x5 필터의 곱으로 만들어지는데, L5S5는 다음과 같이 구해진다.

$$L5^T \times S5 = \begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ -4 & 0 & 8 & 0 & -4 \\ -6 & 0 & 12 & 0 & -6 \\ -4 & 0 & 8 & 0 & -4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix} \quad (1)$$

Color Channel Cb, Cr에 적용되는 평균 필터는

$$M = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} / 16 \quad (2)$$

이다.

이 외에도 두 개의 1x5 필터에 콘볼루션을 적용하면 식 (1)과 유사한 다양한 5x5 텍스처 필터가 얻어진다.

2.2 로스 에너지

2D 로스 필터를 구하면 각 필터는 2D 텍스처 영상과 콘볼루션하여 텍스처 에너지를 계산한다. 영상 I(i, j)가 주어지면, 첫 단계는 조명 효과를 제거하기 위하여 픽셀 주변의 15x15 윈도우의 그레이스케일 값의 평균을 구한 후에, 픽셀로부터 구한 평균을 차감한다. 따라서, 출력 값은 0을 중심으로 분포된다. 다음으로 K개의 2D 로스 필터를 적용하는데, 식 (3)의 F_k(i, j)는 영상 I(i, j)를 k번째 필터 M_k로 콘볼루션하여 얻은 결과이다.

$$F_k(i, j) = I(i, j) \otimes M_k \quad (3)$$

여기서 \otimes 는 콘볼루션 연산자이다. $k \in \{1, \dots, K\}$.

이 과정에서 콘볼루션 연산을 거친 모든 픽셀은 텍스처 에너지

값으로 계산된다. 따라서, k번째 텍스처 에너지 E_k는 식 (4)와 같이 계산된다. 픽셀 주변의 15x15 블록의 에너지의 합이 현재 픽셀의 에너지가 된다.

$$E_k(r, c) = \sum_{i=r-7}^{r+7} \sum_{j=c-7}^{c+7} |F_k(i, j)| \quad (4)$$

여기서 r과 c는 픽셀의 좌표값이다.

3. 자동 깊이 생성 방법

각 텍스처 에너지는 입력영상에 k번째 필터를 적용하여 얻는다. 얻어진 에너지들의 합은 최종적으로 깊이 값으로 변환한다. 에너지의 범위가 [E_{min}, E_{max}]이면, 로스 깊이 D_L은 에너지 E로부터 다음 선형 변환식을 이용하여 구해진다.

$$D_L = D_{min} + \frac{E - E_{min}}{E_{max} - E_{min}} \cdot (D_{max} - D_{min}) \quad (5)$$

여기서 [D_{min}, D_{max}]는 사용자가 지정하는 깊이 값의 범위로, 실험에서 D_{min}=0, D_{max}=255로 하였다.

로스 필터를 콘볼루션하여 얻은 깊이 맵은 모든 픽셀의 깊이 값을 가지고 있다. 그런데, 이 깊이 맵으로부터 전경객체 및 배경을 분리하는 것은 어렵다. 즉, 분할작업은 3D 모델링 데이터를 필요로 한다. 따라서 그림 2의 자동 깊이 생성 방법을 제안한다.

로스 깊이 맵으로부터 특징 점(feature point)을 구하고, 구해진 특징 점에 델러노이 삼각화(Delaunay triangulation)를 적용하여, 삼각형 메쉬(mesh)로 분할한다. 삼각형 각 정점의 깊이 값을 구로 셰이딩(Gouraud shading) 보간 방법을 거쳐 최종 깊이 맵을 생성한다.

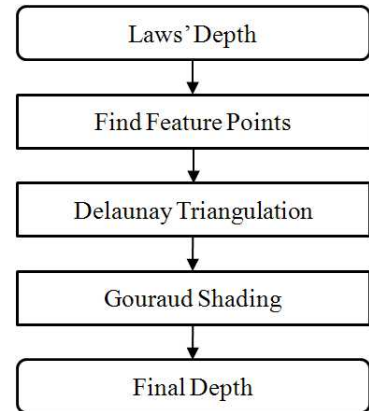


그림 2. 자동 깊이 생성 방법의 개요

특징 점은 일정한 영역에서 깊이 값이 큰 값으로 정하였다. 영상을 그리드로 분할하고 그리드의 교차점을 중심으로 그림 3의 N x N 탐색창 B에서 최대 깊이 값을 갖는 픽셀을 식 (6)을 이용하여 찾는다.

$$(m, n) = \arg_{(m, n) \in B} \max_{(m, n) \in B} D(i - m, j - n) \quad (6)$$

여기서 D(i-m, j-n)은 영상의 (i-m, j-n)에 위치하는 깊이 값이다.

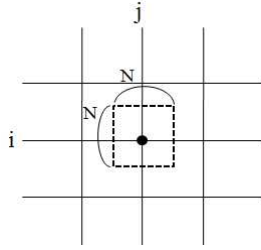


그림 3. search window의 범위

구해진 초기 특징 점에서 중요하지 않은 점을 제거하기 위해서 임계값(threshold)보다 작으면 해당 픽셀을 특징 점에 포함하지 않았다.

텔러노이 삼각화(Delaunay Triangulation) [10]는 임의의 세 점들을 연결하는 외접원을 그리고, 그 외접원이 다른 한 점을 포함하지 않을 경우 이 세 점에 의하여 삼각형이 결정된다. 이렇게 결정된 삼각형들은 삼각형의 가장 작은 각을 최대화 하여 가능한 예각 삼각형이 되는 특징을 갖는다. 전체 영역에서 구해진 좌표를 이용하여 그림 4와 같이 삼각 메쉬를 구성한다.

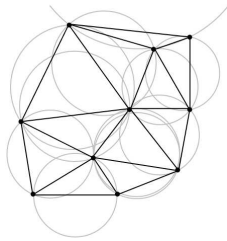


그림 4. 외접하는 삼각형으로 만든 텔러노이 삼각형

텔러노이 삼각형으로 분할된 깊이 맵은 구로 셰이딩을 이용하여 삼각형 내부를 각 삼각형 정점의 깊이 값으로 보간한다. 구로 셰이딩으로 보간된 최종 깊이 맵을 구한다.

4. 실험 결과

실험에서 사용한 영상은 그림 5에서 보여준다. 실험 영상은 객체와 배경이 동시에 존재하는 영상과, 배경만 존재하는 영상으로 구분된다. 실험영상의 해상도는 640x480이다.



그림 5. 실험영상: Butterfly, Lenna, Water, Desert

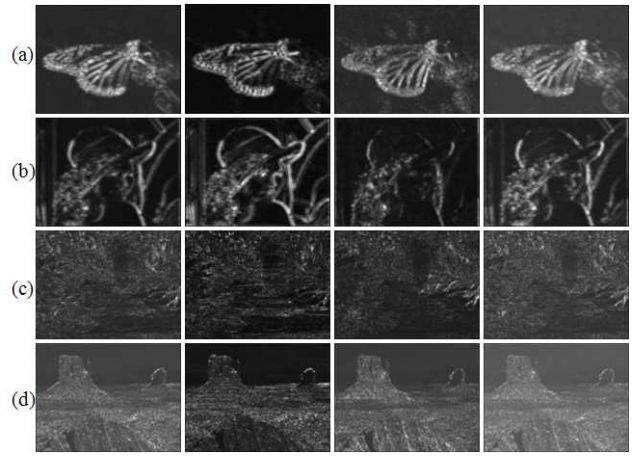


그림 6. 다양한 로스 필터를 적용한 깊이 맵. (a) ESE5, ESL5, ESR5, ESS5, (b) L5E5, L5L5, L5R5, L5S5, (c) R5E5, R5L5, R5R5, R5S5 및 (d) S5E5, S5L5, S5R5, S5S5

그림 6은 실험영상에서 로스 필터를 적용한 결과이다. Butterfly 영상에서는 ESE5와 ESL5필터를 적용하여 객체와 배경이 뚜렷하게 분리되었고, 배경보다 객체에 더 큰 깊이를 얻었다. 배경 영역에서는 큰 깊이나 경계가 나타나지 않았다. Lenna 영상에서는 L5L5를 적용한 영상에서 가장 큰 깊이를 얻었고 얼굴의 윤곽은 검출하였지만 배경과 완벽히 분리하지는 못하였다. L5S5를 적용한 영상에서 보다 부드러운 깊이를 확인하였다. Water 영상은 객체가 없는 배경만 존재하는 영상으로 큰 깊이나 뚜렷한 경계는 나타나지 않았고, R5R5를 적용하여 폭포와 나무의 깊이가 차이나는 것을 확인하였다. Desert 영상은 입력 영상이 가진 거친 텍스처가 모든 필터에서 나타났고, S5L5를 적용한 영상에서 지평선과 바위의 경계가 가장 뚜렷하게 나타났다. 영상 앞쪽의 바위에서 좀 더 큰 깊이를 갖는 것을 확인하였다.

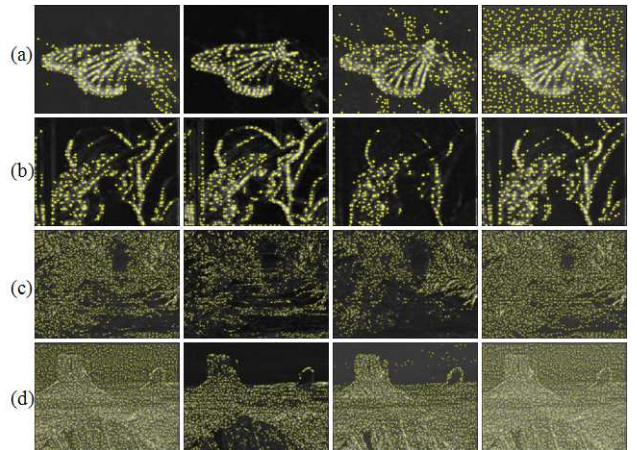


그림 7. 그림 6의 깊이맵에서 얻어진 특징점. (a), (b)의 임계값은 50, (c), (d)의 임계값은 60 임.

그림 7은 그림 6을 그리드로 분할하고, 찾은 각 영역의 특징 점들을 보여준다 (yellow로 표시됨). (a)와 (b)의 특징 점들은 임계값을 50으로 하였고, (c)와 (d)의 임계값은 60으로 하였다. 이 점들에 텔러노이 삼각화를 적용하여 얻어진 메쉬 구조는 그림 8에서 보여준다.

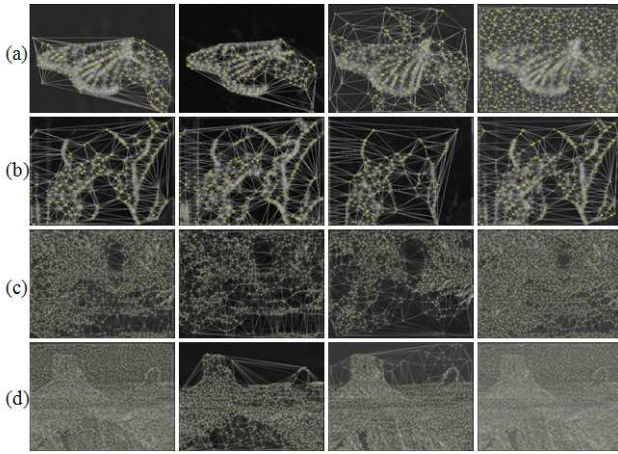


그림 8. 그림 7의 텔러노이 삼각화를 적용한 삼각형 메쉬 구조.

그림 8의 Butterfly 영상에서 E5E5와 E5L5필터를 적용한 객체가 분리된 영상이 비교적 모델링이 잘되었고, Lenna 영상은 L5L5를 사용한 영상이 가장 잘 되었다. Water 영상은 배경 존재하는 영상으로 뚜렷한 객체 모델링이 이루어지지 않았고, 폭포와 나무의 경계만 확인할 수 있었다. Desert 영상도 객체가 없는 영상이지만 S5L5 필터를 적용한 영상에서 앞쪽 바위 부분이 모델링 된 것을 확인하였다.

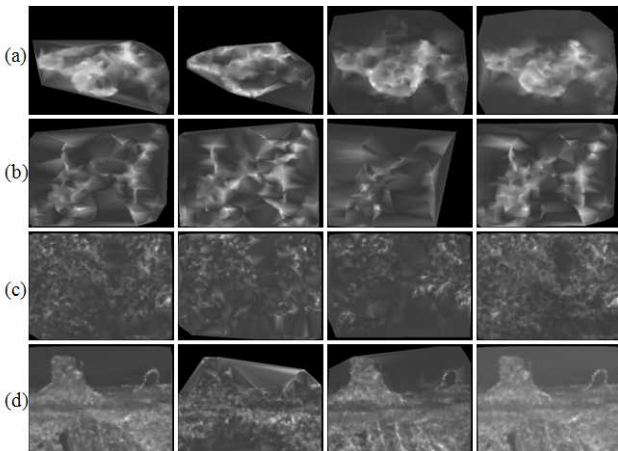


그림 9. 최종적으로 얻어진 깊이 맵.

그림 9는 그림 8에 구로 웨이딩을 적용하여 얻은 최종 깊이 맵이다. Butterfly 영상의 배경 영역에서는 큰 깊이나 경계는 나타나지 않았다. 나머지 영상은 임계값보다 큰 영역의 검출로 인해 배경과 전경을 정확히 분리하지 못하였다. Lenna 영상은 L5L5를 적용한 영상에서 가장 큰 깊이가 나타났고 얼굴의 윤곽은 검출하였지만, 배경과 완벽히 분리하지는 못하였다. Water 영상은 특별한 전경 객체가 존재하지 않아 큰 깊이나 경계를 검출하지 못하였고, 나무 부분에서 일정한 텍스처가 존재하는 것을 확인하였다. Desert 영상은 S5L5와 S5R5를 적용한 영상에서 바위의 경계와 지평선을 경계로 깊이 차이가 나는 것을 확인하였다.

5. 결론

본 논문에서는 2D 영상의 깊이를 얻기 위해 로스 텍스처 필터를 사용하여 자동 깊이 생성 방법을 제안하였다. 텍스처를 분석하기 위해 평균 명암, 에지, 스포트, 리플, 웨이브 등의 특성을 검출하는 필터를

사용하였고, 영상에 필터를 컨볼루션하여 텍스처 에너지를 계산하였다. 텍스처 에너지를 로스 깊이 맵으로 변환하고, 깊이 맵 영상에서 특징 점을 찾아 삼각 메쉬로 분할하였다. 분할된 삼각형 내부를 각 정점의 깊이 값으로 보간하여 최종 깊이 맵을 얻었다. 배경과 객체가 존재하는 영상에서 배경과 분리된 객체를 확인할 수 있었고, 특징 점을 찾는 과정에서 정확하지 못한 텍스처 때문에 삼각 메쉬를 통한 모델링의 어려움이 있었다. 최종적으로 얻어진 깊이 맵도 영상 텍스처의 특성만을 가지고 생성한 깊이 정보이기 때문에 정교하지 못한 깊이를 얻었다.

향후 연구로는 특징 점 선택 시 보다 정교한 특징 점 선택과, 다양한 텍스처 필터 연구를 통해 보다 개선된 깊이 맵 생성 방법 연구가 필요하다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음. (NIPA-2009-(C1090-0902-0017))

참고 문헌

- [1] A. Redert, et al., "ATTEST : Advanced Three- Dimensional Television System Technologies", First International Symposium on 3D Data Processing Visualization and Transmission, pp. 313-319, Jun. 2002.
- [2] Wang, S, "Depth from Shading Based on 2D Maximum Entropy", Intelligent Computation Technology and Automation (ICICTA), 2008 International Conference on Volume 2, 20-22 Oct. 2008 Page(s):119 - 121
- [3] Chao-Chung Cheng; Chung-Te Li; Po-Sen Huang; Tsung-Kai Lin; Yi-Min Tsai; Liang-Gee Chen; "A block-based 2D-to-3D conversion system with bilateral filter", Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference on 10-14 Jan. 2009 Page(s):1 - 2
- [4] Murino, V.; Vernazza, G.; Santos-Victor, J.; Sentieiro, J.; "Depth estimation by an adaptive depth from motion algorithm", OCEANS '94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation.' Proceedings, Volume 1, 13-16 Sept. 1994 Page(s):1/193 - 1/198 vol.1
- [5] Simon, C.; Bicking, F.; Simon, T.; "Depth estimation based on thick oriented edges in images", Industrial Electronics, 2004 IEEE International Symposium on Volume 1, 4-7 May 2004 Page(s):135 - 140 vol. 1
- [6] D. F. McAllister (editor), Stereo computer graphics and other true 3D technologies, Princeton, NJ:Princeton University Press, 1993.
- [7] E. R. Davies, Laws' texture energy in Texture. In Machine Vision: Theory, Algorithms, Practicalities, 2nd Ed., Academic Press, San Diego, 1997.
- [8] Motofumi T. Suzuki, Yoshitomo Yaginuma and Haruo Kodama, "A 2D Texture Image Retrieval Technique Based On Texture Energy Filters", IMAGAPP 2009 - International Conference on Imaging Theory and Applications, pp. 145-151.
- [9] J. Michels, A. Saxena, and A. Y. Ng, "High speed obstacle avoidance using monocular vision and reinforcement learning", Proc. of International Conference on Machine Learning, ACM, 2005.
- [10] O. Cheong, M. V. Kreveld, M. Overmars, Computational Geometry: Algorithms and Applications, Springer-Verlag, 2008.