

셀프 캘리브레이션을 위한 컬러 스테레오 이미지의 대응 관계 추출

김도윤, 정명진
한국과학기술원 전자전산학과

Color stereo image matching for self-calibration

Do-Yoon Kim and Myung Jin Chung
Dept. of EECS, Korea Advanced Institute of Science and Technology

Abstract - 본 연구에서는 캘리브레이션 박스와 같은 사전에 약속된 물체를 사용하지 않고 일반적인 환경 (unstructured environment)의 컬러 스테레오 영상으로부터 특징점을 찾고, 특징점 사이의 대응관계를 사용자 개입 없이 파악하는 방법에 대해 소개한다. 또한 찾은 대응관계를 이용해 스테레오 카메라 사이의 에피폴라 기하학(epipolar geometry) 관계를 계산하여 셀프 캘리브레이션에 이용한다. 이와 유사한 연구는 많이 진행되어 왔으나 대부분의 연구가 흑백 영상에서 진행되어 왔다. 본 연구에서는 컬러 이미지의 속성을 이용해 흑백 영상을 이용할 때보다 외부 환경의 변화에 강인하며, 정밀한 대응 관계를 찾을 수 있음을 실험을 통해 보인다.



그림 1 흑백 영상에서 harris 방법을 이용한 모서리 검출

1. 서 론

주변 환경을 파악하기 위해 사용되는 여러 센서 중에서 가장 많은 정보를 전달해 주는 센서는 영상 정보를 취득할 수 있는 카메라이다. 카메라를 통해 얻은 영상 정보에는 색상(color), 형태(shape)가 존재하며, 여러 장의 이미지나 두 대 이상의 카메라를 사용할 경우에는 거리(depth), 움직임(motion) 정보를 추가로 얻을 수 있다. 이러한 장점으로 인해 로봇 연구자와 컴퓨터 비전 연구자들은 영상으로부터 얻고자 하는 정보를 추출하기 위한 많은 연구를 시도하고 있다.

영상 처리 연구는 대부분 특징점(feature point)으로부터 시작된다. 일반적으로 1차원 정보를 담고 있는 경계선(edge)나 2차원 정보를 담고 있는 모서리(corner)가 특징점으로 많이 사용되는데, 이 중 모서리는 경계선에 비해 노이즈에 강인한 특성이 있어 스테레오 매칭(stereo matching)이나 물체의 추적(tracking)등에 많이 사용된다. 모서리가 이미지의 특징점으로 유용하기 때문에 영상에서 모서리를 빠르고 노이즈에 강인하게 검출하려는 시도는 계속되고 있다[1][2][5].

현재까지 알려진 모서리 검출법은 Harris 방법, SUSAN 방법, CSS 방법들이 있다. 이 중 Harris 방법은 밝기 영상의 미분값으로 자기상관행렬을 만든 후, 이 행렬의 고유값 차이가 작은 곳은 모서리로 정의하는 방법으로 다른 방법에 비해 영상의 회전, 스케일 변화에 강인하게 모서리가 검출되어 가장 많이 사용된다[1][2]. Harris 모서리 검출법을 간단히 살펴보면 다음과 같다. 영상 밝기 값을 $I(x, y)$ 라 정의하고, 영상의 가로 방향을 x , 세로 방향을 y 라고 놓을 때, 영상의 각 위치에서 2×2 행렬 M 은 다음과 같이 정의된다.

$$M = \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \quad (식 1)$$

M 행렬의 고유값(eigenvalues) λ_1, λ_2 에서 값이 작은 고유값이 사용자가 정한 한계값(threshold) t 보다 크면 그 점을 모서리로 정의한다.

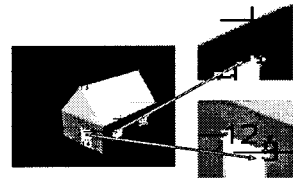


그림 2 흑백 영상에서 대응점 계산 시 발생하는 문제

이러한 방법들은 모두 이진 영상(binary image)이나 흑백 영상(gray image)을 대상 이미지로 하기 때문에 밝기 변화가 심하거나 밝기의 차이로는 두드러지지 않는 특징점들을 검출하기가 힘들었다. 또한 두 이미지 사이에서 모서리를 특징점으로 대응 관계를 계산하는 경우 영상의 정보에서 밝기값 만을 사용하기 때문에 잘못된 대응 관계로 귀결되는 결과가 많다[2]. 이러한 단점은 사용자의 개입 없이 자동으로 대응관계를 추출해야 하는 셀프 캘리브레이션에서는 심각한 문제를 만들어낸다[8]. 본 논문에서는 Harris 방법을 확장하여 컬러 스테레오 이미지로부터 특징점을 추출하는 방법에 대해 제시한다. 추출된 특징점은 컬러 정보를 이용해 두 영상의 대응 관계를 계산하는데 이용된다. 이 때 기존에 많이 사용되는 RGB 컬러 공간의 단점을 살펴보고 RGB 컬러 공간에 비해 많은 장점을 가지고 있는 CIELAB 컬러 공간에 대해 살펴본다. 이 때 RGB 컬러 공간에서 CIELAB 컬러 공간으로 변환하는데 많은 시간이 걸리는 기존의 방법을 lookup 테이블을 통해 실시간으로 변환하는 방법도 함께 제시한다. 계산된 대응 관계로부터 fundamental 행렬 F 를 계산하고 기존 방법에 비해 정밀도가 많이 향상됨을 보인다. 또한 제안한 방법으로 구한 F 는 셀프 캘리브레이션을 통해 3차원 복원 문제에 쉽게 적용할 수 있음을 보인다.

본 논문은 다음과 같이 구성되어 있다. 2 장에서는 컬러 영상에서 모서리 검출 방법을 설명하고, 검출된 모서리를 이용해 두 장의 영상에서 대응점을 계산하는 방법에 대해 제안한다. 3 장에서는 제안한 방법의 실험과 기존의 방법에 의한 결과를 비교하여 제안한 방법의 타당성을 조사해 본 후, 4 장에서 결론을 맺는다.

2. 컬러 특징점 추출 및 대응

2.1 컬러 특징점의 정의

흑백 영상에서 사용하던 방법을 컬러 영상에 적용할 때 가장 흔히 사용되는 방법은 컬러 모델- RGB 또는 HSI 등의 각 채널에 흑백 영상에서 사용된 방법을 사용한 후 각 채널의 결과를 더하는 방식을 사용하고 있다 [2]. 하지만 RGB 공간이나 HSI 등 일반적으로 잘 알려진 컬러 모델은 각 채널의 스케일이 같지가 않다. 다시 말하면 선형 공간이 아니어서 '+'와 같은 연산이 성립되지 않는다. 따라서 청색이 많은 영상과 노란색이 많은 영상에서는 상대적으로 노란색이 많은 영상에서 많은 특징점이 나오게 되며 이러한 문제는 단순한 weigh 로서 해결되지 않는다 반면 CIELAB 으로 명명된 컬러 공간은 각 채널이 각각 L, A, B 라는 이름으로 불리는데, 이 채널의 스케일은 선형적인 관계를 가지고 있어 '+'의 연산 개념이 성립된다. 따라서 흑백 영상에서 장점이 많이 알려진 연산자들을 컬러 영상에 적용하는데 CIELAB 공간은 많은 장점을 가지고 있어 최근 많이 연구되고 있다 [3]. 하지만 카메라로부터 입력되는 신호를 직접 CIELAB 컬러 모델의 형식으로 받아들일 수 없기 때문에 RGB 컬러 모델에서 정의된 영상 정보를 CIELAB 컬러 모델의 형식으로 바꿀 필요가 있다.

2.1.1 RGB 컬러 공간에서 CIELAB 컬러 공간으로의 변환

RGB 공간에서 CIELAB 공간으로 바꾸는 일반식은 다음과 같이 두 단계를 거쳐 변환된다.

단계 1) 장치에 독립적인 XYZ 공간으로 변환한다.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{식 2})$$

단계 2) XYZ 공간에서 CIELAB 공간으로 변환한다.

$$\begin{aligned} L' &= 116 \times (Y/Y_n)^{1/3} \text{ with } Y/Y_n > 0.008856 \\ L' &= 903.3 \times Y/Y_n \text{ with } Y/Y_n \leq 0.008856 \\ a &= 500 \times (f(X/X_n) - f(Y/Y_n)) \\ b &= 200 \times (f(Y/Y_n) - f(Z/Z_n)) \\ \text{where } f(t) &= t^{1/3} \text{ with } t > 0.008856, \\ f(t) &= 7.787 \times t + 16/116 \text{ with } t \leq 0.008856 \end{aligned} \quad (\text{식 3})$$

이 변화는 영상의 한 점마다 곱셈 연산이 9번, 덧셈 연산이 6번이 사용되기 때문에 영상의 크기가 320x240 일 경우 PC에서 127,795,200번의 머신 사이클이 필요하며 P-III 1GHz의 프로세서를 사용하는 경우 2초 정도의 시간이 필요하다. 따라서 실시간 시스템으로 사용하기 위해서는 빠른 변환이 필요한데 이러한 변환식의 복잡도로 인해 CIELAB 컬러 공간은 사용되지 못했다.

2.1.1 LUT를 이용한 빠른 CIELAB 컬러 공간 변환

Lookup table(LUT)를 이용한 빠른 변환 방법은 off-line 과정과 on-line 과정으로 나뉜다. 입력 영상은 RGB 컬러 모델에서 16비트 형식을 사용한다. 이 때 상위 5비트는 R 채널에 해당하며 중간의 6비트는 G 채널, 하위 5비트는 B에 해당한다.

off-line 과정) 각각의 값을 8비트로 변환한 후, 식 2와 식 3을 통해 각 컬러값에 해당하는 CIELAB 값을 테이블로 만든다. 테이블은 각각 65Kbyte의 크기를 가지게 된다. 이 때 5비트(또는 6비트)의 값을 8비트로 확장하는데 필연적으로 오차가 발생하게 되는데, linear approximation을 통해 가장 가까운 값을 지정한다.

on-line 과정) off-line 과정에서 만들어진 LUT는 index값이 입력 영상의 컬러값이고, 값은 각각 L, A, B 채널의 값이 된다. 따라서 on-line 과정에서는 단순

한 배열의 참조만으로 CIELAB 컬러 공간의 값을 얻을 수 있다.



그림 3 off-line으로 RGB2CIELAB 변환 과정
그림 4 off-line으로 만들어진 lookup table을 통해 CIELAB 영상으로 바꾸는 과정

	기존 방법	제안한 방법
계산 시간 (P-III 1GHz 320x240 이미지)	2s	off-line)2.5s on-line)1us
각 채널당 error	0%	L(0.11%). A(0.32%). B(0.33%)

표 1. 기존 방법과 제안한 방법의 비교

표 1에서 볼 수 있듯이 실제 사용시에는 on-line의 변환 시간만 소요되므로 실시간 시스템에 응용이 가능하며, 각 채널 당 에러는 0.4% 이하로 문제가 되지 않는다.

2.2 SVD를 사용한 스테레오 매칭

추출된 특징점(본 논문에서는 모서리)으로부터 두 장의 입력 영상 사이의 대응점 관계를 알기 위해서는 에피폴라 기하학을 알아야 한다. 하지만 에피폴라 기하학을 계산하기 위해서는 영상의 대응점을 알아야 하는 모순이 생긴다. 이런 문제점은 두 영상 사이의 이동이 작다는 가정 하에 특징점을 추적하거나 옵티컬 플로우 등을 통해 매칭 문제를 해결하곤 했다 [6][7]. 하지만 일반적인 스테레오 영상의 경우 무시할 수 없을 정도의 이동(베이스라인)과 각 카메라의 회전 이동이 존재하므로 이러한 방법으로는 일반적인 영상의 매칭 문제를 해결할 수 없다.

따라서 본 논문에서는 흑백 영상의 매칭 문제를 해결하는데 사용하던 SVD 스테레오 매칭 해결 방법 [7]을 변형하여 컬러 특징점의 대응 관계를 파악하는데 사용한다. 두 장의 입력 영상을 각각 I, J 라고 가정하면, 2.1 절에서 제안한 CIELAB 컬러 공간으로 두 영상을 변환한 후, 각 채널에 harris 방법으로 모서리를 검출한다. 이 때 검출된 모서리는 각각 m, n 개라 놓고, 각 영상에서 I_i, J_j 로 표시한다. 이 때 각 특징점 사이의 정규화된 상관관계는 다음의 식으로 표현할 수 있다.

$$C_{ij}^{ch} = \frac{\sum_{m=1}^M \sum_{n=1}^N (A_m - \bar{A})(B_m - \bar{B})}{W^2 \sigma(A) \sigma(B)} \quad (\text{식 4})$$

단, W 는 주어진 특징점을 중심으로 사용자가 지정한 블록의 크기이며, \bar{A}, \bar{B} 는 해당 블록의 평균값이다. 또한 ch 인덱스는 각 컬러 채널을 의미한다. 따라서 두 특징점이 완벽히 같은 컬러 정보를 가지고 있다면 각 채널에서 C_{ij} 값은 1이 되며, 서로 전혀 다른 정보를 가지고 있으면 -1의 값을 가지게 된다.

이러한 특징점 사이의 상관관계만으로는 반복되는 형태나 일반 환경에서 특징점 사이의 관계를 구명하는데 많은 문제점을 보인다 [7]. 따라서 특징점 사이의 거리 정보를 추가하여 적절한 거리 안에서 특징점이 서로 매칭이 되도록 한다.

위의 내용으로 새롭게 정의되는 특징점 사이의 상관 관계값은 다음과 같다. 이 때 컬러 공간에서 L 채널값은 사용하지 않는다. L 채널을 밝기 정보를 가지게 되는데, 입력 영상에서 컬러 정보는 상대적으로 밝기(intensity) 정보에 비해 덜 영향을 받으므로, 밝기 정

보에 해당하는 L 채널의 정보는 사용하지 않고, 컬러 정보만 가지고 있는 A, B 채널만을 사용하여 특징점 사이의 상관 관계값을 계산해 낸다.

$$G_{ij} = \frac{C^A_{ij} + C^B_{ij} + 2}{4} e^{-r_{ij}/2\sigma} \quad (\text{식 } 5)$$

식 5를 통해 G_{ij} 를 구하면 $m \times n$ 행렬 G 가 구해진다. G 는 singular value decomposition(SVD)을 통해 다음과 같이 분리된다.

$$G = TDU^T \quad (\text{식 } 6)$$

식 6에서 D 행렬을 대각 행렬은 1로 설정되고, 나머지 요소들은 0으로 지정되어 있는 행렬 E 로 대치하여 식 7을 통해 새로운 행렬 G_{new} 를 구한다.

$$G_{new} = TEU^T \quad (\text{식 } 7)$$

G_{new} 행렬은 각 매칭점 사이에 상관 관계를 나타내는 행렬로 1:1 매핑이 가능하게 된다. G_{new} 행렬로부터 적절한 임계값을 사용하여 매칭점을 추출한다.

3. 실험

3.1 컬러 특징점 추출 및 에피폴라 기하학 관계 계산

2장에서 제안한 방법으로 컬러 스테레오 영상에서 특징점을 추출, 매칭점을 찾은 후 이를 이용해 fundamental 행렬 F 를 계산한 후, 에피폴라 기하학을 통해 구한 F 의 정밀도를 조사해 보았다.



그림 5 일반 환경에서 컬러 특징점 추출

그림 6에서 볼 수 있듯이 구한 F 를 사용하여 왼쪽 영상에서 한 점을 지정하면 그 점에 해당하는 에피폴라 라인을 왼쪽 영상에 그렸는데, 대응점 위에 에피폴라 라인이 존재하는 것을 확인해 볼 수 있다.

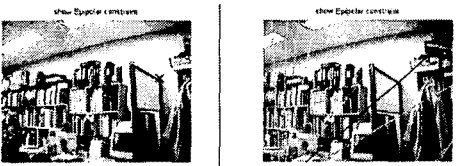


그림 6 매칭으로부터 구한 에피폴라 기하학을 이미지에 표시

3.2 셀프 캘리브레이션에 적용

3.1에서 구한 F 로부터 카메라의 파라미터를 추출하는 셀프 캘리브레이션을 수행하면, 스테레오 영상으로부터 3차원 복원이 가능하다. 셀프 캘리브레이션은 Kruppa 식으로부터 내부 파라미터를 복원하는 방법, F 수식에서 직접 카메라 파라미터를 복원하는 방법, kalman filter를 사용하여 여러 장의 이미지로부터 수행하는 방법 등이 소개되어 왔다(9). 이 중 Brooks에 의해 제안된 direct method는 스테레오 카메라가 평행하게 설치되어 있는 경우 내부 파라미터와 외부 파라미터를 F 로부터 직접 구할 수 있는 방법이다. 본 논문에서는 3.1절에서 구한 F 로부터 brooks가 제안한 방법으로 카메라 파라미터를 구한 후, 입력 영상에 존재하는 물체를 3차원 복원하여 그 정밀도를 살펴본다.

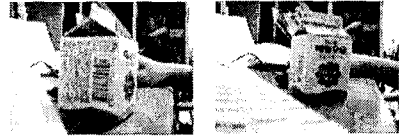


그림 7 스테레오 입력 영상과 3차원 복원된 영상(아래)

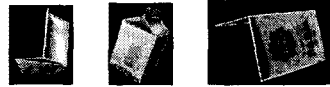


그림 7에서 볼 수 있듯이, 3차원 복원은 스케일을 제외한 나머지 파라미터로 복원이 가능했으며, euclidean 공간에서 90도로 표현되는 부분이 복원된 결과에서도 91도로 표현되는 것을 확인할 수 있었다.

4. 결 론

스테레오 영상에서 매칭점을 파악하는 것은 매우 중요한 문제로써 많은 연구자들에 의해 연구가 되어 왔다. 최근 셀프 캘리브레이션에 대한 관심이 높아지면서 매칭점을 영상 노이즈에 대해 강인하고, 자동으로 추출해야 할 필요성이 높아졌는데, 본 논문에서는 흑백 정보만이 아닌 컬러 정보를 이용하여 그 정도를 높였다. 따라서 특별한 패턴이 필요하지 않고 일반적인 환경으로부터 특징점을 추출하는 것이 가능했으며, SVD 매칭 방법을 통해 두 영상의 fundamental 행렬을 구하는 것이 가능하다. 구해진 F 행렬로부터 직접 캘리브레이션이 가능한 다이렉트 방법을 통해 카메라 파라미터를 구한 후 3D 복원을 통해 구한 F 의 정밀도가 우수함을 보였다. 본 논문에서 제안한 방법은 능동형 스테레오 카메라나, 이동형 로봇, 인간형 로봇 등에 적용된 카메라의 캘리브레이션에 유용하게 사용될 수 있을 것으로 기대된다.

(참고 문헌)

- [1] C. Harris, M. Stephens, "A combined corner and edge detector," Int. Proc. of the 4th Alvey Vision Conference, pp 147-151, 1988
- [2] P. Montesinos, V. Gouet, R. Deriche, "Differential Invariants for Color Image", Int. Conf. on Pattern Recognition(ICPR98), 1998
- [3] Y.S Moon and I.S. Kweon, "Robust Dominant Motion Estimation Algorithms against Local Linear Illumination Variations", Technical Report of Robotics and Computer Vision Lab. Dept. of EECS KAIST, 1999
- [4] C. Schmid, R. Hohn, and C. Bauckhage, "Evaluation of Interest Point Detectors," INRIA Technical Report
- [5] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," Pattern Recognition Letters, 1:95-102, 1982
- [6] R. Jain, R. Kasturi and B.G. Schunck, "Machine Vision," McGraw-Hill, 1995
- [7] Maurizio Pilu, "A Direct Method for Stereo Correspondence based on Singular Value Decomposition," Int. Proc. in CVPR, pp261-266, 1997
- [8] P. F. McLauchlan and D. W. Murray, "Active Camera Calibration for a Head-Eye Platform using the Variable State-Dimension Filter," Int. Jour. of IEEE PAMI, Vol. 18, No. 1, pp15-22, 1996
- [9] M.J. Brooks, L. de Agapito, D.Q. Huynh and L. Baumela, "Direct Methods for self-calibration of a moving stereo head", Int. proc. of the 4th European Conference on Computer Vision, Lecture Notes pp 415-426. 1996