

에피폴라 기하를 이용한 특징점 정합†

권혁민*, 한준희* 정연구**

* 포항공과대학교 전자계산학과 컴퓨터비전 연구실

** 한국전자통신연구원 삼성공학연구부

Feature Point Matching using Epipolar Geometry

H. M. Kwon* J. H. Han* Y. K. Jeong**

* Computer Vision Lab, Dept. of Comp. Sci. & Eng., POSTECH

** Sensitivity Engineering Research Dept., ETRI

요 약

본 논문은 두 장의 스테레오 영상으로부터 자동적으로 특징점 정합을 수행하도록 하는 한 방법을 제안한다. Correlation기반의 특징점 정합을 빠르고 안정적으로 수행하며 이 때에 발생하는 애매성 문제에 대한 해결방법을 제시한다. 또한, LMedS방법을 사용하여 outlier를 효과적으로 제거시키고 에피폴라 기하를 이용하여 정합의 성능을 향상시킨다. 실내, 실외 영상에 대한 다양한 실험결과를 본 논문에서 제안하는 방법이 빠르고 효율적임을 보여준다.

1. 개요

“특징점 정합(feature point matching)”이란 한 쪽 영상에서의 어떤 한 특징점에 대응되는 다른 쪽 영상 안에서의 한 점을 찾는 작업을 말한다(이를 correspondence problem이라 한다). 다시 말해서, image1안의 한 특징점 p 에 대응되는 image2에서의 특징점 q 를 찾는 작업이 바로 특징점 정합이다. 이러한 정합의 문제는 컴퓨터 비전분야에서 오랫동안 대두되어 왔고, 가장 중요한 문제중의 하나이다. 특히, 스테레오 비전, 샵프-켈리브레이션, 3차원 복원과 같은 분야에 있어서 특징점 정합은 상당히 중요한 위치를 차지한다.

본 논문에서는 특징점 정합을 수행하기 위하여 다음과 같은 절차를 사용한다.

- 1 특징점의 추출: 각 영상으로부터 특징점들을 뽑아낸다.
- 2 초기 정합: 각 영상에 해당하는 특징점집합 사이에 초기정합을 수행한다.
- 3 애매성의 제거: 초기정합의 결과로 발생하는 애매성을 제거한다.
- 4 Outlier의 제거: 잘못된 정합된 정합쌍들을 제거한다.
- 5 추가 정합: 정합되지 못한 특징점들을 다시 한 번 검토한다.

기존의 방법들은 자연 영상에 부적합한 직선을 정합에 이용하거나[6] 초기 정합에 너무 많은 제약조건을 사용하여 애매성을 오히려 증가시키기도 하고[3], 애매성 해결을 위해 relaxation기법을 이용함으로써 정합의 속도를 많이

떨어뜨린다는 등의 단점을 갖는다. 본 논문은 안정적이고 빠른 정합에 그 주안점을 둔다. 전체 프로시저를 구성하는 각 단계가 이 목적을 위해 선택 개발되었고, 여러 스테레오 영상에 대한 실험결과를 이를 입증한다. 또한, 전체 과정이 완전히 자동적으로 수행되도록 하였다.

2. 특징점의 추출

영상안에서 특징점이란 코너점과 같이 높은 곡률값을 갖는 점으로서, 이 점들은 영상안에서 다른 점들과 구별되어지는 점들이다. 특징점의 추출을 위해 Smith와 Brady가 SUSAN corner detector[2]라는 것을 제안하였는데, Harris가 제안한 Plessey detector[1]가 방향성 미분값에 기초한 방법인데 반해, 이 추출기는 밝기값 비교에 기초한 방법이다.

영상안에서 한 점 $p(x,y)$ (nucleus)와 그 점을 중심으로 하는 마스크 윈도우 W 를 고려해보자. W 안에서의 다른 한 점을 $q(x,y)$ 라 정하고, 또, $I(p)$, $I(q)$ 가 각각 점 p , q 에서의 밝기값을 나타낸다고 하자. 그러면 p 와 q 사이의 밝기값 유사도 $S(p,q)$ 는 다음과 같이 정의된다.

$$S(p,q) = \begin{cases} 1 & \text{if } |I(p) - I(q)| < t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

여기에서 t 는 밝기값 유사정도를 나타내는 임계값이다. 이제 점 p 에서의 “USAN(Unvalue Segment Assimilating Nucleus)”는 다음과 같이 계산된다.

$$N(p) = \sum_W S(p,q) \quad (2)$$

이 값이 마스크 윈도우 면적의 반보다 작으면 그 점 p 는 특징점으로 볼 수 있고, 또 이 값이 작을수록 특징점의 정도가 커지게 된다. 이 SUSAN 특징점 추출기는 무엇보다도 그 수행속도가 빠르다는 장점을 갖는다. 또한, 이 추출기는 좋은 반복성(repeatability)을 가지므로 영상으로부터 특징점들을 뽑아내는데 적합하다.

† 본 연구는 한국과학기술연구원(KIST)의 한국전자통신연구원(ETRI)으로부터 연구비의 일부를 지원받았음

3. Correlation에 의한 초기 정합

특징점 추출기를 사용하여 각 영상으로부터 자동적으로 특징점 집합(feature point set)을 얻어냈다면, 그 다음의 절차는 이 특징점 집합들간에 초기 정합을 실시하는 일이다. 초기 정합을 수행하는 데에는 SSD(Sum of Squared Difference), NCC(Normalized Correlation Coefficient), Zabih의 rank transform[4]과 같은 방법을 사용할 수 있는데 본 논문에서는 다른 방법들에 비해 보다 안정적인 NCC 방법을 사용하였다.

NCC방법에서의 상관계수를 구하는 방법을 살펴보자. 두 영상에서의 특징점을 각각 $p_1(x, y), p_2(x, y)$ 라 하자. 그리고, 이 점들을 중심으로하는 $m \times n$ 마스크 윈도우를 고려하자. 그러면, $p_1(x, y), p_2(x, y)$ 사이의 상관계수는 다음과 같이 계산된다.

$$C(p_1, p_2) = \frac{1}{K} \sum_{u=-m}^m \sum_{v=-n}^n [I_1(p_1.x + u, p_1.y + v) - \bar{I}(p_1)] \times [I_2(p_2.x + u, p_2.y + v) - \bar{I}(p_2)] \quad (3)$$

여기에서 $K = (2m+1)(2n+1)\sigma(p_1)\sigma(p_2)$ 이며, \bar{I} 와 σ 는 각각 괄호안의 점을 중심으로 하는 마스크 윈도우내에서의 평균 밝기값과 표준편차를 나타낸다. 또한 위 지수는 -1에서 1까지의 범위를 가지게 된다.

4. 주변 특징점의 정합 결과를 이용한 애매성의 제거

초기 정합의 과정에서, 첫 번째 영상에서의 한 특징점 p 에 대한 대응점 q 를 찾기위해 p 와 두 번째 영상에서의 정합 가능한 모든 특징점 사이에 NCC 값을 조사하게 되는데, 이 때 가장 큰 NCC 값을 나타내는 두 번째 영상에서의 점 q 를 p 의 대응점이라고 판단할 수는 없다. 왜냐하면, 다른 특징점들이 참의 특징점보다 얼마든지 더 높은 상관계수를 가질 수 있기 때문이다. 따라서, 특징의 NCC값을 임계값으로 설정하여 그 이상의 값을 갖는 두 번째 영상에서의 점 q 들을 모두 p 와 대응되는 점이라고 판단하게 되는데, 그렇게 하면 한 점 p 에 다수의 점 q 가 대응되는 애매성(ambiguity)의 문제가 발생하게 된다. 이 애매성의 문제를 효율적으로 해결하기 위하여 주변 특징점들의 정합 결과를 이용한다. 그것은 정합이 잘 된 정합쌍의 주변에는 역시 다수의 정합쌍들을 발견할 수 있다는 논리로부터 나온다.

먼저, 두 특징점 p, q 사이의 정합 정도를 $MS(p, q)$ 라 표기하자. 그러면 본 논문은 $MS(p, q)$ 의 계산을 위해 다음의 식을 제안한다.

$$MS(p, q) = C_{pq} \sum_{r \in \mathcal{N}(p)} \left[\frac{\text{unique}_{s \in \mathcal{N}(q)} C_{rs}}{\text{dist}(p, q; r, s)} \times \frac{\text{max_direct} - \text{direct}(p, q; r, s)}{\text{max_direct}} \right] \quad (4)$$

여기에서,

$$\begin{aligned} \text{dist}(p, q; r, s) &= \frac{1}{2}[d(p, r) + d(q, s)] \\ \text{direct}(p, q; r, s) &= |(p.x - r.x) - (q.x - s.x) \\ &\quad + (p.y - r.y) - (q.y - s.y)| \\ \text{max_direct} &= 4 \times (\text{radius of neighbor}) \end{aligned}$$

C_{pq}, C_{rs} 는 각각 p, q 와 r, s 사이의 상관계수를 나타내며, $\text{dist}()$ 는 p, q 로부터 멀리 떨어져 있는 이웃 정합쌍일수록 그

영향력을 줄이기위해 사용한다. 그리고, $\text{direct}()$ 는 그 값이 작을 수록 이웃 정합쌍이 정합쌍 p, q 와 상대적으로 같은 방향성을 가짐을 나타낸다. 또한, $MS(p, q)$ 의 계산식에서 unique 연산자는 이웃 특징점의 중복 계산을 피하기 위해 사용된다 즉, p 의 주변 특징점들이 q 의 주변 특징점들과 “다대일” 또는 “다대다”의 관계를 가질 수 있기 때문에 정합쌍 p, q 에 대해 상대적으로 최대의 정합 정도를 갖는 유일한 이웃 정합쌍만을 $MS(p, q)$ 의 계산에 사용하기위해 이 unique 연산자를 사용한다.

5. Outlier의 제거

특징점 정합에 있어서 outlier란 초기 정합의 결과 중 잘못 정합된 특징점 순서쌍을 말한다. 따라서, 이 outlier들을 제거하는 과정은 특징점 정합의 수행에서 반드시 필요한 과정이다. 널리 알려진 outlier 제거 알고리즘으로는 M-estimator[5]와 LMedS[3]방법이 존재한다. 그러나, 본 논문에서는 잡음에 대해 덜 민감한 LMedS(Least Median of Squares) 방법을 사용한다.

초기 정합과 애매성 제거를 마친 후 갖게되는 정합쌍의 갯수를 n 이라 하자. Fundamental matrix F 를 구하는데는 최소한 8개의 정합쌍이 필요하므로 n 개의 정합 결과 중 임의로 $p = 8$ 개를 선택한다. (이렇게 n 개의 결과중 임의로 8개를 선택하는 것을 J 로 색인시킨다.)

그러면, 특징의 J 번째 집합으로부터 얻어지는 fundamental matrix F_J 에 대한 “median of squared residuals” M_J 는 다음과 같이 결정된다.

$$M_J = \text{med}_{i=1, \dots, n} [d^2(\mathbf{m}_{2i}, F_J \mathbf{m}_{1i}) + d^2(\mathbf{m}_{1i}, F_J^T \mathbf{m}_{2i})] \quad (5)$$

그리고, 모든 M_J 중에서 최소의 값을 갖는 M_J 를 SM_J 로 놓는다. 이제 robust standard deviation을 다음의 식을 이용하여 계산한다.

$$\hat{\sigma} = 1.4826[1 + 5/(n - p)]\sqrt{SM_J} \quad (6)$$

마지막으로, 정합결과 중 outlier들은 다음의 식을 이용하여 찾는다

$$\text{outlier} = \begin{cases} \text{True} & \text{if } r_i^2 \leq (2.5\hat{\sigma})^2 \\ \text{False} & \text{otherwise} \end{cases} \quad (7)$$

여기에서 $r_i^2 = d^2(\mathbf{m}_{2i}, F \mathbf{m}_{1i}) + d^2(\mathbf{m}_{1i}, F^T \mathbf{m}_{2i})$ 이다.

6. 에피폴라 라인에 의한 추가정합

이상의 과정을 거치면 얻어진 정합쌍으로부터 에피폴라 기하(epipolar geometry)를 구할 수 있다. 단순히 에피폴라 기하만 구하는 것이 목적이란 outlier를 제거하는 과정까지만 수행하면 되지만, 보다 많은 정합쌍이 요구되는 상황(예를 들어, 셀프-캘리브레이션이나 3차원 복원 등등)이라면 미처 정합되지 못한 특징점들에 대해 다시 한번 검토해보는 것이 바람직하다.

Outlier를 제거시키고 나면 상당히 정확한 fundamental matrix를 얻을 수 있고, 따라서 두 특징점 집합간에 에피폴라 라인(epipolar line)을 찾을 수 있다. 이 에피폴라 라인은 첫 번째 영상의 한 특징점 p 에 대응될 수 있는 두 번째 영상에서의 점 q 를 찾는 데 또 하나의 제약조건으로 작용한다. 결국, 에피폴라 라인을 중심으로 하는 좁은 밴드에 걸쳐 있는(단, 이미 정합쌍으로 판정된 특징점은 제외시킨다) 특징점에 대해서만 그 정합 가능성을 살펴보면 된다.

본 논문에서는 outlier 제거 후 얻어지는 평균 에피폴라 라인 오차(epipolar line error)만큼의 오차 허용치에 걸쳐있는 특징점만을 그 정합대상으로 보았다. 이렇게 하면, 추가 정합쌍이 적어진다는 단점이 있으나 기존의 에피폴라 라인 오차는 악화시키지 않게 된다.

7. 실험 결과

실험은 다양한 실내, 실외 영상에 대하여 수행되었다. 이들 실험을 통하여 제안하는 방법의 특징점 정합이 빠르고 효율적임을 보인다. 먼저, 그림 1은 mosaicbox 영상에 대해 특징점을 뽑아낸 결과이다. 그리고, 그림 2는 이 영상들에 대해 특징점 정합을 수행한 결과이다.



그림 1. 좌우 mosaicbox 영상에 대한 특징점 추출결과

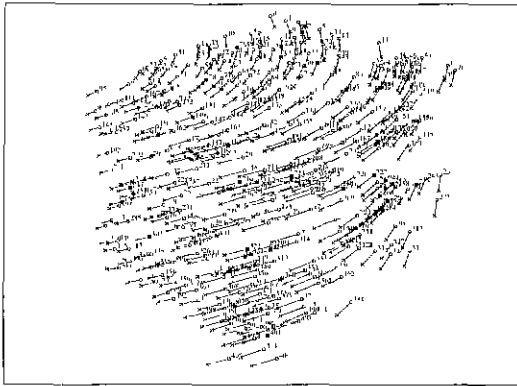


그림 2: mosaicbox 영상에 대한 특징점 정합결과

그림 3는 각각 school 영상과 pillow 영상에 대하여 본 논문의 알고리즘을 적용시켰을 때의 특징점 정합결과를 나타낸다. 두 영상 모두 정합이 잘 이루어졌음을 알 수 있다.

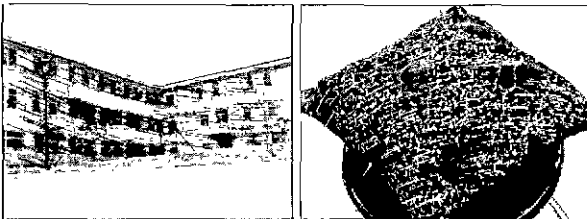


그림 3: school 영상과 pillow 영상에 대한 특징점 정합결과

특징점 정합이 잘 이루어졌는지에 대한 좋은 판단기준은 정합결과로 얻은 정합 순서쌍으로부터 fundamental matrix를 구한 다음 이 행렬을 다시 정합결과에 적용하여 평균 에피폴라 라인 오차를 살펴보는 것이다. 표 1은 위에서 실험한 각각의 영상에 대해 정합 단계에서마다의 에피폴라 라인 오차와 정합 순서쌍의 갯수를 나타낸다.

영상	2단계 후	3단계 후	최종 오차	정합 수
mosaicbox	5.46412	0.945795	0.313175	407(6)
school	47.8696	4.80227	0.499917	252(11)
pillow	10.7714	4.31724	0.290295	267(14)

표 1: 정합 단계에서마다의 평균 에피폴라 라인 오차와 정합 수

정합 수 항목에서 괄호안의 숫자는 정합의 마지막 단계에서 추가로 얻어진 정합쌍의 수를 나타낸다. 추가된 정합쌍의 갯수가 적는데 이는 앞에서도 언급했듯이 에피폴라 라인의 오차 허용치를 작게 했기 때문이다. 또한 실험한 바에 의하면, 에매성 제거단계를 거쳤을 때의 오차가 초기정합시 최대의 NCC값을 갖는 점으로 대응점을 결정했을 때의 최종 오차보다 대부분 좋았으며, 특징점의 정합이 상당히 빠르게 수행되었다(한 예로, 좌우 각각 1800여개, 1700여개의 특징점을 갖는 pillow영상에 대한 특징점 정합에서 SGI O2 미신으로 11초 정도가 소요된다).

8. 결론

지금까지 두 장의 스테레오 영상에 대해 특징점 정합을 수행하는 과정과 그에 대한 실험을 살펴보았다. 각 영상으로부터 특징점들이 SUSAN 특징점 추출기에 의해 빠르게 추출되며 제안하는 방법의 correlation 역시 빠르고 안정적으로 수행된다. 초기 정합시 발생하는 에매성의 제거와 LMedS에 의한 outlier의 제거는 정합의 정확성을 훨씬 높여주며 에피폴라 기하를 이용한 추가정합 과정은 정합쌍의 수를 늘려준다. 아울러, 제시하는 방법의 일련 과정이 완전히 자동적으로 수행된다.

여러가지의 실험 예제를 웹페이지

<http://falcon.postech.ac.kr/ipawb/demo/> 에서 자세히 볼 수 있다.

참고문헌

- [1] C. G. Harris, "Determination of Ego-motion from Matched Points," *Proc. Alvey Vision Conf.*, 1987.
- [2] S. M. Smith and J. M. Brady, "A New Approach to Low Level Image Processing," *Int J Comput. Vision*, vol. 23(1), pp. 45-78, 1997.
- [3] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras, "Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig," *ECCV*, vol. 800, pp. 567-575, 1994.
- [4] Ramn Zabih and John Woodfill, "Non-parametric Local Transforms for Computing Visual Correspondence," *ECCV*, vol. 801, pp. 151-158, 1994.
- [5] Soren I Olsen, "Epipolar Line Estimation," *ECCV*, pp. 307-311, 1992.
- [6] P. Beardsley, P. Torr and A. Zisserman, "3D Model Acquisition From Extended Image Sequences," *ECCV*, pp. 683-695, 1996.