

논문 2004-41SC-4-5

능동적인 비전 시스템에서 카메라의 시선 조정: 컴퓨터 비전과 제어의 융합 테마

(Steering Gaze of a Camera in an Active Vision System: Fusion Theme of Computer Vision and Control)

한 영 모*

(Young mo Han)

요 약

능동적인 비전 시스템의 전형적인 한 테마는 카메라의 시선 고정 문제이다. 여기서 카메라의 시선 고정이란 동적인 물체 상의 지정된 한 점이 항상 이미지의 중앙부에 놓이도록 카메라의 자세를 조정하는 것으로서, 이를 위해서는 카메라에 비친 영상 정보를 분석하는 기능과 카메라의 자세를 제어하는 두 가지 기능이 결합되어야 한다. 본 논문에서는 영상분석과 자세제어가 한 개의 프레임 하에서 설계되는 카메라의 시선 고정을 위한 알고리즘을 제안한다. 이 때 제작시의 어려움을 피하고 실시간 응용을 위해서 본 알고리즘은 카메라의 calibration이나 3차원 거리 정보의 복원을 필요로 하지 않도록, 그리고 닫힌 형태(closed-form)가 되도록 설계된다.

Abstract

A typical theme of active vision systems is gaze-fixing of a camera. Here gaze-fixing of a camera means by steering orientation of a camera so that a given point on the object is always at the center of the image. For this we need to combine a function to analyze image data and a function to control orientation of a camera. This paper presents an algorithm for gaze-fixing of a camera where image analysis and orientation control are designed in a frame. At this time, for avoiding difficulties in implementing and aiming for real-time applications we design the algorithm to be a simple closed-form without using any information related to calibration of the camera or structure estimation.

Keywords: active vision, gaze-fixing, orientation control..

I. 서 론

현재 사용되고 있는 비전 시스템(vision system)의 성능 향상을 위해서 종종 카메라(camera)의 시선 제어(gaze control) 기술이 요구된다. 카메라의 시선 제어는 소위 방향 조절이 가능한 비전 시스템(directed vision system)의 형태로 구현될 수 있다. 방향 조절이 가능한

비전 시스템은 종종 능동적인 비전 시스템(active vision system)으로 불리운다. 전통적인 스테레오 비전(stereo vision)과는 달리, 능동적인 비전은 원하는 정보를 얻을 수 있도록 현재의 영상정보를 분석한 후, 그 결과를 활용해 카메라의 자세를 조정한다.

능동적인 비전의 초기 연구는 Tenebaum에 의해 행해졌는데, "영상의 부적합성(image inadequacy)"이라는 용어를 사용해서 3차원 정보 복원 문제에서 정적인 비전 시스템(static vision system)의 한계를 지적하였다^[1] 이와 같은 견지에서 많은 학자들이 능동적인 비전 시스템이 정적인 비전 시스템에 비해 우수한 많은 장점을 가지고 있다는 점을 지적하였다^{[2][3]}. 예를 들어, 카메

* 정회원, 이화여자대학교 정보통신공학과
(Dept. of Information Electronics Ewha Womans University)

※ 이 논문은 2004년도 두뇌한국21사업에 의하여 지원되었음.

접수일자: 2004년4월20일, 수정완료일: 2004년7월7일

라의 이동은 시야(field of view)를 크게 신장 시킬 수 있다. 또한 능동적인 비전은 중심과 비전(fovea vision)을 구현 가능케 하는데, 중심과 비전이란 가장 좋은 영상을 제공할 수 있는 카메라 렌즈 중심 부분(일명 중심와)에 영상이 향시적으로 맺히도록 하기 위해 카메라 방향을 조절하는 시스템이다. 따라서 대상체의 관심 부위를 적은 오차로 카메라의 중심 부분에 머무르게 하는 것이 가능하다.

위와 같은 목적으로 능동적인 비전 시스템은 많은 학자들에 의해 연구되고 구현되었다^{[4][5][6][7][8]}. 기존에 알려진 많은 알고리즘의 특징은 발견적인 형태(heuristic form)이거나, 카메라의 calibration 정보를 사용하거나, 또는 3차원 거리 정보를 복원하기 위해 다수의 카메라를 사용하는 경우가 많았다. 하지만 발견적인 형태의 알고리즘은 실시간 응용에 어려움이 있는데, 그 이유는 성능예측이 어렵고, 심지어 경우에 따라서는 많은 계산량이 필요하게 될 수 있기 때문이다. 한편 카메라의 calibration과 복수의 카메라 시스템은 구현 시 상당한 어려움에 부딪힐 수 있는데, 그 이유는 카메라의 calibration 알고리즘이 복잡하고, 복수의 카메라 영상들 사이에 존재하는 일치점(correspondence)을 정확히 찾기가 어려울 뿐 아니라 카메라들 간의 정보융합(data fusion)의 문제가 발생하기 때문이다.

위와 같은 기존 연구에서 발생하는 어려움을 극복하고 실시간 응용을 목표로, 본 논문에서는 카메라의 calibration 정보나 3차원 거리 정보를 사용하지 않고, 또한 한 개의 카메라로 구현 가능한 닫힌형태(closed-form)의 알고리즘을 제안한다.

II. 카메라의 calibration 정보와 3차원 거리 정보를 사용하지 않는 능동적인 비전 시스템

그림 1에서 묘사된 카메라의 제어 시스템을 생각해 보자. 공간좌표계 {SF}는 공간상에 고정된 좌표계로서, 원점이 카메라의 렌즈 중심에 있고, k-축은 이점에서 렌즈에 수직하게 놓여 있다. 카메라는 공간좌표계 {SF}의 원점을 중심으로 자유롭게 회전할 수 있으며 카메라 좌표계 {CF}가 부착되어 있다. 편의상 카메라 좌표계 {CF}의 원점과 공간좌표계 {SF}의 원점이 일치하도록 놓여있다. 또한 이미지(image) 평면은 렌즈 중심으로부터 카메라 좌표계 {CF}의 k-축을 따라 초점거리(focal length) f 의 거리에 있다. 이 이미지 평면에는 이미지

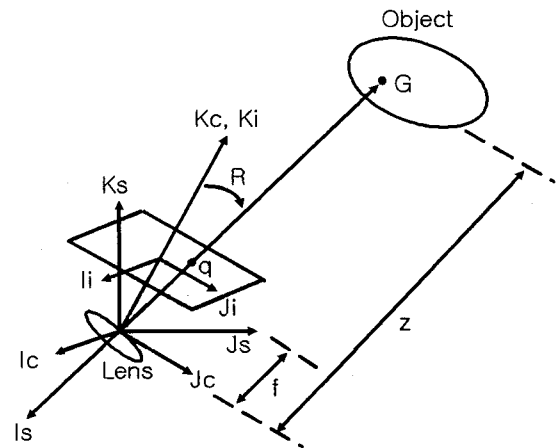


그림 1. 능동적인 비전 시스템
Fig. 1. Active Vision System.

좌표계 {IF}가 부착되어 있는데, 각 축은 카메라 좌표계 {CF}의 각 축과 같은 방향으로 놓여 있다.

주어진 강체(rigid body) 상의 한 점 $G \in R^3$ 이 경사 투영(perspective projection)에 의해 이미지 상에 q 점으로 투영되었다고 하자. 이 경우 다음의 관계식이 성립한다.

$$G = (x, y, z)^T$$

$$q = f(G/z) = f(x/z, y/z, 1)$$

여기서 x, y, z 는 각각 카메라 좌표계 {CF}에서 바라본 G 점의 i, j, k 축 방향 성분이다.

이제부터 우리가 원하는 것은 G 점이 항상 카메라의 시선(line-of-sight) 상에 위치하게 하는 것이다. 다시 말해서 q 점이 항상 이미지 좌표계 {IF}의 원점 부분에 오게 하고자 한다. 본 논문에서는 편의상 G 점을 시선고정점(gaze-fixing point)이라 부르기로 한다. 사실상 시선고정점은 설계자의 목적에 따라 임의로 선택될 수 있는데, 예를 들어 대상체에 부착된 운동좌표계(body-fixed frame)의 원점을 시선고정점으로 선택할 수 있다.

이제 우리의 목표는 대상체가 정지해 있든 또는 운동을 하든 상관없이, 항상 q 점이 이미지 좌표계 {IF}의 원점에 머무르도록 카메라의 자세(orientation)를 조정하는 효과적인 방법을 찾아내는 것이다.

카메라를 회전행렬 $R \in SO(3)$ 만큼 회전해서 q 점이 이미지 좌표계 {IF}의 중심에 놓이게 되었다고 하면 다음의 관계가 성립한다.

$$R\hat{z} = \hat{G} \quad (1)$$

여기서

$$SO(3) = \{R \in R^{3 \times 3} | R^T R = I, \det(R) = 1\}$$

$\dot{z} = [0, 0, 1]^T$ 은 카메라 좌표계 {CF}에서 바라 본 이미지 좌표계{IF}의 z-축 방향의 단위 벡터이고, G 의 좌표가 카메라 좌표계 {CF}에서 바라 본 값이며, $\hat{G} = G/||G||$ 이다.

$G = zq/f$ 의 관계로부터 식(1)은 다음과 같이 쓰여질 수 있다.

$$R\dot{z} = \hat{q} \tag{2}$$

여기서 $\hat{q} = q/||q||$.

그럼 q 점이 항상 이미지 좌표계{IF}의 중심에 오게 하려면 식(2)를 만족하는 회전행렬 $R \in SO(3)$ 을 찾으면 된다. 이 문제 정의에서 주목할 점은 카메라의 calibration 자료 f 와 3차원 거리 정보 z 를 사용하지 않는다는 것이다.

III. 단위 구(Unit Sphere)상에서 정의된 시선제어(Gaze Control)

q 가 식(2)를 만족하게 하는 회전행렬 $R \in SO(3)$ 을 찾기 위해서, 다음과 같은 점근적인(asymptotic) 제어 시스템을 설계하고자 한다.

$$R(t)\dot{z} \rightarrow \hat{q} \text{ as } t \rightarrow \infty$$

이때 $R(t)\dot{z}$ 은 크기가 항상 1이므로 사실상 단위 구 $S^2 = \{\Phi \in R^3 | ||\Phi|| = 1\}$ 의 원소이다. 따라서 본 제어 시스템 설계 문제는 실제적으로 단위 구상에서의 문제로 이해될 수 있다. 그리고 이러한 단위 구상의 문제 정의가 카메라의 calibration 정보와 3차원 거리 정보의 사용을 배제할 수 있는 방법을 제공한다고 볼 수 있다.

S^2 상에서 정의된 점근적인 제어 시스템을 설계하기 위해서 먼저 상태변수(state variable) X 와 제어변수(control variable) U 를 아래와 같이 정의하자.

$$X = R\dot{z} \in S^2 \tag{3}$$

$$U = \dot{X} \in T_X(S^2) \tag{4}$$

여기서 $T_X(S^2)$ 는 X 에서의 S^2 의 접평면 공간(tangential space)를 나타낸다.

그럼, $X(t) \rightarrow \hat{q}$ as $t \rightarrow \infty$ 을 만족하는 제어는 아래

에 주어진 제안 1의 형태로 정리될 수 있다.

제안 1: 식(3)과 식(4)에 의해 표현되는 제어 시스템을 고려하자. $X(0)$ 와 $\hat{q}(0)$ 사이의 초기 각도가 $\pi/2$ 보다 작다고 가정하자. 그러면, 상태 변수 $X(t) \in S^2$ 가 점근적으로 동적인 목표점 $\hat{q}(t)$ 에 다가가게 하는 제어 $U = U(X(t))$ 는 다음과 같이 주어진다.

$$U = (I - XX^T)(\dot{\hat{q}} + k\hat{q}) / (X^T \dot{\hat{q}}) \tag{5}$$

여기서 $k \in R$ 은 임의의 양의 상수이다.

증명: $T_X(S^2)$ 에서 바라 본 $X(t)$ 부터 $\hat{q}(t)$ 까지의 거리를 $E(t)$ 라 하면 다음과 같은 값을 갖는다.

$$E(t) = (I - X(t)X(t)^T)(\hat{q}(t) - X(t)) \in T_X(S^2) \tag{6}$$

다음으로 $E(t)$ 의 제곱크기를 생각해 보자

$$W = \frac{1}{2} ||E||^2 \tag{7}$$

이 값을 시간 t 에 대해 미분하면 다음식을 얻는다.

$$\dot{W} = \{(I - XX^T)\dot{\hat{q}}\} - \{\dot{X}X^T + X\dot{X}^T\}\hat{q} + (I - XX^T)\dot{\hat{q}} \tag{8}$$

관계식 $(I - XX^T)^T = (I - XX^T)$ 과 $(I - XX^T)X = 0$ 을 활용하여 식(7)을 다음과 같이 단순화시킬 수 있다.

$$\dot{W} = \{(I - XX^T)\dot{\hat{q}}\} - \{\dot{X}X^T\}\hat{q} + (I - XX^T)\dot{\hat{q}} \tag{9}$$

식(4)와 식(5)를 식(9)에 대입함으로써 다음의 관계식을 얻는다.

$$\dot{W} = -kW \tag{10}$$

따라서 $W(t) \rightarrow 0$ as $t \rightarrow \infty$. 또한 $\theta \in R$ 이 X 와 \hat{q} 사이의 각도라 할 때, $||E|| = |\sin\theta|$ 의 관계가 있음을 보일 수 있다. 그리고 $0 \leq \theta \leq \pi/2$ 의 범위에서 $\sin\theta$ 는 전단사, 단조 증가 함수이므로, $X = \hat{q}$ 일 때만 $W = 0$ 이다. 따라서 다음과 같은 특성을 갖게 된다.

$$X(t) \rightarrow \hat{q} \text{ as } t \rightarrow \infty.$$



식(5)는 투영된 이미지가 q 인 대상체 상의 한 점을 카메라가 점근적으로 따라가도록 조종하는 알고리즘의

한 예이다.

마지막으로 제안 1에서 사용된 \hat{q} 를 일반적인 측정값 q , \dot{q} 로부터 구하는 방법에 대해 살펴보자.

따름정리 1: 임의의 벡터 $V \in R^n$ 에 대해서 다음의 관계식이 성립한다^[11].

$$\hat{V} = \frac{1}{\|V\|} (I - \hat{V}\hat{V}^T) \dot{V} \quad (11)$$

증명: 직접적인 계산을 통해 증명할 수 있다. ■

따름정리 1의 결과를 사용해서 \hat{q} 를 다음과 같이 계산할 수 있다.

$$\hat{q} = \frac{1}{\|q\|} (I - \hat{q}\hat{q}^T) \dot{q} \quad (12)$$

IV. 성능 평가와 검증

성능 평가를 위해 본 논문에서 제안하는 카메라의 시선 제어 알고리즘을 기존의 선형제어 방법과 비교해 보자. 이 선형 제어 방법은 참고문헌 [10]에서 기술된 시각적 추적 시스템(visual tracking system)의 모델을 선형화 한 후 PID 제어 법칙을 적용한 것이다(제어이득 상수는 참고문헌^[12]에서 제시된 방법을 사용하였다). 여기서 성능 평가 지수로는 추적 에러(tracking error), 즉 현재 카메라의 방향과 목표 방향 사이의 각도를 선택하였다. 초기 추적 에러가 0.1585이고 2 픽셀 잡음(2-pixel noise)를 가정했을 때의 추적 에러는 그림 2에 주어져 있다. 그리고 본 모의실험(simulation) 중의 추적 에러와 계산시간의 평균치는 표 1에 주어져 있다.

그림 1과 표 1를 살펴보면, 본 논문에서 제안한 방법이 선형 제어 방법에 비해 추적에러가 상당히 적음을 알 수 있다. 그 이유는 다음과 같다. 선형 제어 방법의 6추적 에러는 크게 두 가지로 구성되는데, 하나는 잡음(noise)에 의한 것이고 또 다른 하나는 모델의 선형화에서 비롯된 것이다. 그리고 후자의 에러가 전자의 것보다 상당히 크다. 이와 같은 선형 제어 방법의 경우와는 대조적으로 본 논문에서 제안한 방법에서는 추적 에러의 요인이 잡음(noise)에 의한 것 하나 뿐이다.

표 1은 본 논문에서 제안한 방법의 경우가 선형 제어

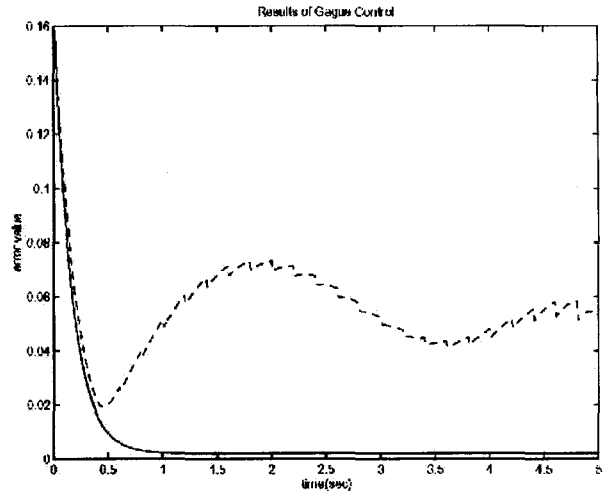


그림 2. 픽셀 잡음이 있는 경우의 시선 제어 결과. X축과 Y축은 각각 시간과 추적에러(현재의 카메라 방향과 목표 방향 사이의 각도)를 나타낸다. 이 모의실험에서 초기 추적에러를 0.1585로 설정하였다. 그래프에서 실선과 점선은 각각 본 논문에서 제안한 방법과 선형 제어 방법의 경우를 나타낸다.

Fig. 2. Results of Gaze Control in the existence of 2-pixel noise. X and Y - axes denote the time and tracking error which is defined as the angle between current and desired orientations of the camera. In this simulation we assume initial tracking error of 0.1585. The solid and the dotted lines denote the cases of our and the linear control laws, respectively.

표 1. 추적에러와 계산시간의 평균치
Table 1. Average value of tracking error and computation time.

	tracking error	computation time
ours	0.0019	0.2410
linear control law	0.0567	0.4903

방법의 경우보다 계산 시간 또한 상당히 적음을 보여주고 있다. 그 이유는 다음과 같이 생각될 수 있다. 선형 제어 방법의 경우 제어 알고리즘을 계산하기 위해서는 두 가지 단계를 거쳐야 하는데, 선형 모델의 계산과 선형제어 알고리즘의 계산이 그것이다. 반면 본 논문에서 제안한 방법의 경우는 제어 알고리즘의 계산이라는 한 가지 계산 단계를 가지고 있다. 뿐만 아니라 본 논문에서 제안한 방법은 선형 제어 방법에 비해 단순한 닫힌 구조 형태(closed-form)로 이루어져 있다.

IV. 결 론

본 논문에서는 이미지 좌표와 이미지 속도가 측정 가능할 때, 대상체 상의 한 점에 카메라의 시선을 점근적으로 고정할 수 있는 제어 알고리즘을 제안하였다. 본 제어 방법은 다음의 두 가지 장점을 가지고 있다. 첫째로 카메라의 calibration이나 3차원 정보 복원 과정이 필요 없다. 둘째로 제어 방법이 매우 쉽고 빨라서 실시간 사용(real-time processing)이 가능하다는 것이다.

참 고 문 헌

- [1] J. Tenenbaum, Accommodation in Computer Vision. Ph. D. Dissertation, Stanford, 1970.
- [2] J. Aloimonos, I. Weiss, and Bandopadhyay, "Active vision," International Journal on Computer Vision, vol. 1, pp. 333-356, 1988.
- [3] D. Ballard, "Animate vision," Artificial Intelligence, vol. 48, pp. 57-86, 1991.
- [4] V. Ortmann and R. Eckmiller, "Real-time object recognition based on active vision and sequential analysis," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 3325-3328, 1999.
- [5] X. Roca, J. Vitria, M. Vanrell, and J. J. Villanueva, "Gaze control in a binocular robot systems," in Proc. of IEEE Conf. on Emerging Technologies and Factory Automation, pp. 479-485, 1999.
- [6] Mario Sznajder, Turner Inanc, and Octavia Camps, "Robust controller design for active vision systems," in Proc. of American Conf. on Control, pp. 2013-2017, June 2000.
- [7] Elsayed E. Hemayed, Mounen T. Ahmed, and Aly A. Farag, "CardEye: A 3D Trinocular Active Vision System," in Proc. of IEEE Conf. on Intelligent Transportation Systems, pp. 39-403, October 1-3, 2000.
- [8] M. Pellkofer and E. D. Dickmanns, "EMS-Vision: Gaze Control in Autonomous Vehicles," in Proc. of IEEE Symposium on Intelligent Vehicles, pp. 296-301, October 3-5, 2000.
- [9] R. H. S. Carpenter, The Eye Moments, Pion Limited, London, 1988.
- [10] N. P. Papanikolopoulos, P. K. Khosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision," IEEE Trans. Robotics and Automation, vol. 9, no. 1, pp. 14-35, 1993.
- [11] M. Bosse, "Plane Clustering from Spherical Image Sequences," MIT, December 1998.
- [12] Gene F. Franklin, J. David, and Abbas Emami-Naeini, Feedback Control of Dynamic Systems," Addison-Wesley, 1991.

저 자 소 개



한 영 모(정회원)

1992년 서울대학교 물리교육학과
학사 졸업.

1995년 서울대학교 제어계측공학과
석사 졸업.

1998년 서울대학교 전기공학부
석사 졸업.

2002년 서울대학교 기계항공공학부 박사 졸업

<주관심분야: 컴퓨터비전, 제어>

