

Optical Flow를 이용한 3차원 운동 정보에 관한 연구 (3D Motion Estimation Using Optical Flow)

조혜리*, 이정무**, 이상욱*

* 서울대학교 전기공학부

** 홍익대학교 전파공학과

요약

운동(motion) 벡터는 보고 있는 카메라와 관측되는 대상물 사이의 상대적인 움직임에 의해서 발생하는 3차원 물체의 속도가 2차원 영상에 투사되어 맺히는 영상에서의 2차원 속도 벡터를 가리킨다. 영상에서 물체의 움직임은 3차원 공간상의 운동을 알 수 있는 중요한 정보로써 물체를 추적하는데 응용되고 있다. 본 논문에서는 여러 장의 연속적인 2차원 밝기 영상으로부터 카메라의 움직임을 추정하는 문제를 다룬다. 기존의 특징 기반 추적 기법에서는 저 단계의 영상 처리 과정에서 모델과 배경의 특징점이 서로 분리되지 않거나, 모델의 특징(feature)이 소실되었을 경우, 추적이 용이하지 못하고, 카메라와 3차원 물체의 병진과 회전 운동에 의해 발생한 움직임의 경우 3차원 표적 특징이 많이 사라져서 오차가 많이 누적되기도 한다.

본 논문에서는 이러한 문제를 해결하기 위하여 목표물 및 배경 특징들을 사용하여 카메라의 운동 정보를 찾아내는 기법을 제안한다. 제안하는 3차원 카메라의 운동 정보 추정 기법은 크게 두 장의 연속된 영상으로부터 3차원 모델과 배경의 많은 특징들에 대한 광류(optical flow) 검색 과정과, 이로부터 취득한 움직임 벡터와 카메라의 비선형 운동 방정식과 Lagrange multiplier를 통한 카메라의 운동 정보 추정 과정으로 구성된다.

Lagrange multiplier를 도입하는 방법을 제안하였다. 기존의 3차원 운동 추정 방법들은 회전 행렬 R 을 선형화하기 위해 회전인자가 아주 작다는 가정을 하고 있지만, 제안하는 알고리즘은 회전인자를 선형화하지 않아서 회전이 큰 경우에도 정확한 값을 구해낼 수 있다. 또한 정확한 운동 정보를 추정해 낸 후 3차원 모델이 주어지면 영상 정합을 기반으로 한 추적 기법에도 응용될 수 있다.

I. 서 론

운동 정보의 추정 문제는 컴퓨터 비전 분야에서 가장 기본적인 문제의 하나로 알려져 있다. 3차원 운동 정보의 추정은 과거에도 많이 연구해 온 분야로서 Longuet-Higgins[1], Tsai and Huang[2], Faugeras[3] 등이 강체의 운동 정보는 단일한 카메라에서 찍은 두 장의 연속적인 영상으로부터 점의 일치로 얻을 수 있는 선형방정식으로 모델링되고 이를 Singular Value Decomposition(SVD)을 사용하여 구할 수 있다는 것을 증명했다.

여러 장의 연속적인 영상에서 물체와 카메라 사이의 상대적인 운동을 측정하는 방법에는 영상에서 3차원 물체의 특징에 대응되는 2차원 특징 즉, corner point나 평면의 경계를 사용하는 방법이 있다[4]. 그러나, 이러한 특징에 기반을 둔 운동 정보 추정은 저 단계 처리 과정에서 모델과 배경의 특징점이 서로 분리되지 않거나, 모델의 특징(feature)이 소실되는 경우가 발생해서 오차가 많이 누적되기도 한다.

본 논문에서는 특징에 기반을 둔 운동 정보 추정의 단점을 보완하기 위해서 3차원 모델과 배경의 많은 특징들을 이용하는 optical flow를 구한 후, 이로부터 얻어진 속도 정보와 카메라와 3차원 물체로부터 비선형 방정식으로 모델링하였다. 그리고 이러한 비선형 방정식의 해를 효율적으로 구하기 위해서

II. Optical Flow

평면 위의 점들의 3차원 물체 속도가 2차원 영상에 투사되어 맺히는 영상에서의 2차원 속도 벡터를 밝기 영상의 공간적·시간적 패턴으로부터 근사화하여 구하는 것이 optical flow의 목적이다.

일반적으로 optical flow를 구하는 방법은 영상의 밝기 값을 미분하는 방법, 영역을 기반으로 하는 정합, 에너지를 기반으로 하는 방법, 위상을 기반으로 하는 방법 등이 있다. 이 방법들은 각각 차이가 있지만 일반적으로 다음과 같은 공통적인 세 단계를 거친다[5].

1. SNR을 향상시키거나 신호의 필요한 영역을 추출하기 위해서 low-pass filter나 band-pass filter를 통해서 filtering을 한다.
2. 공간적·시간적 미분이나 국부적 상관관계와 같은 기본적인 측정량을 추출한다.

3. 2차원 flow field를 만들기 위해서 측정량을 하나로 합친다.

본 논문에서는 영상의 밝기는 공간과 시간 영역에서 연속적이고 미분 가능하고, 움직이는 물체의 밝기는 일정하다라는 가정을 통해서 다음과 같이 영상의 밝기 값을 미분하는 방법으로 optical flow를 구하였다.

시간에 따른 영상의 밝기를 $E=E(x,y,t)$ 라고 하면 움직이는 물체의 밝기는 일정하기 때문에 식 (1)과 같은 영상 밝기 불변 방정식(image brightness constancy equation)을 이끌어 낼 수 있다.

$$-aE(x(t),y(t),t) = \frac{\partial E}{\partial x} \frac{ax}{dt} + \frac{\partial E}{\partial y} \frac{ay}{dt} + \frac{\partial E}{\partial t} = 0 \quad (1)$$

$$E_x u + E_y v + E_t = 0$$

식 (1)을 일반적으로 optical flow 조건 방정식이라고 한다. 여기서 u, v 는 각각 optical flow의 x 축 y 축에 대한 성분이다. 그러나 식 (1)은 u, v 에 대한 부정방정식이기 때문에 다른 조건이 필요하다.

본 논문에서는 Lucas와 Kanade[6]가 구한 optical flow의 방법을 사용하였다. Lucas와 Kanade는 인접해 있는 화소의 그룹에 대해서 optical flow를 사용하고, 인접해 있는 화소의 그룹은 같은 속도를 가진다는 가정 하에서 선형 시스템을 설계하였다. 각각 작은 공간 Ω 에 인접해 있는 v 에 대하여 식 (1)에 맞는 식 (2)를 최소화하는 v 를 weighted least-square를 사용하여 풀었다.

$$\sum_{x \in \Omega} W^2(x) [\nabla I(x, t) \cdot v + I_t(x, t)]^2 \quad (2)$$

여기서 $W(x)$ 는 Gaussian window함수이다. 이 방법은 식 (3)과 같은 2×2 선형 시스템의 해를 구하는 것과 관련된다.

$$\sum_{x,y} W(x,y) I_x I_x u + \sum_{x,y} W(x,y) I_y^2 v = - \sum_{x,y} W(x,y) I_y I_t$$

$$\sum_{x,y} W(x,y) I_x^2 u + \sum_{x,y} W(x,y) I_x I_y v = - \sum_{x,y} W(x,y) I_x I_t \quad (3)$$

위의 방법으로 구한 optical flow의 결과는 그림 1과 같다.

III. 카메라 모델

카메라를 통해서 영상 위에 3차원 공간상의 점을 투사시키는 과정은 내부인자와 외부인자에 의해서 결정된다. 전형적인 외부인자는 world coordinate에 의한 병진과 회전운동이다. 한편 내부인자는 초점 거리, 줌, 시계와 같은 카메라 인자와 샘플링 비율, 중형비, image offset과 같은 디지털화 과정에서 발생하는 인자가 있다. 광학적인 왜곡을 무시하면, 실제적인 pin-hole 카메라 모델은 3차원 좌표 $X = (X, Y, Z)^T$ 와 이 삼차원 좌표가 영상에 투영된 점 $x = (x, y)^T$ 를 homogeneous coordinate를 사용하여 다음 식 (4)와 같이 표현될 수 있다.

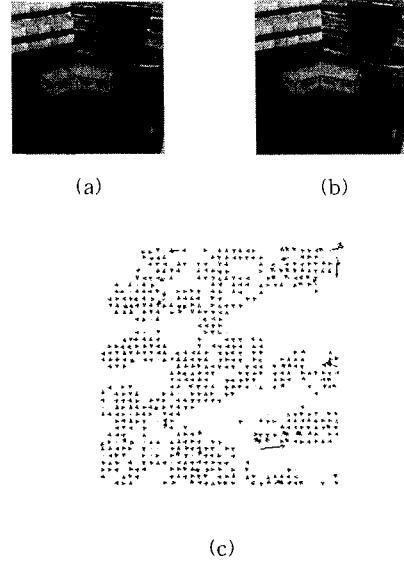


그림 1. 두 장의 연속된 합성영상 (a), (b) 에서 구한 optical flow (c).

$$sm = P[R, t] M = P[RX + t] \quad (4)$$

$$\text{단, } m=[x^T, 1]^T, M=[X^T, 1]^T, P=\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R=\begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix}, \quad t=[t_x, t_y, t_z]^T, \quad r_i=[r_{i1}, r_{i2}, r_{i3}]^T \quad i=1,2,3$$

회전인자이고, t_x, t_y, t_z 는 병진인자이다. 그리고 s 는 영상에서 3차원 물체까지의 거리(depth)를 가리킨다.

두 장의 연속된 영상이 있을 때, 첫 번째 영상을 기준 영상으로 보고 주어진 영상 위의 점을 r_0 라고 하면 식(4)에 의해서 depth에 따른 물체의 좌표계를 다음과 같이 나타낼 수 있다.

$$X(s) = R_0^{-1} [s P^{-1} m_0 - t_0] \quad (5)$$

단, $m_0=[r_0^T, 1]^T$ 이다[3].

두 번째 영상에 대응되는 $r_1(s)$ 는 식 (6)과 같이 표현되어 진다.

$$r_1(s) = [U_1(s)/s_1(s), V_1(s)/s_1(s)] \quad (6)$$

여기서 기준 영상의 $R_0=I$ 로 보면

$$\begin{aligned} & [U_1(s), V_1(s), s_1(s)]^T \\ &= P[R_1 X(s) + t_1] \\ &= P[R_1 R_0^{-1} (s P^{-1} m_0 - t_0) + t_1] \\ &= P[R(s) P^{-1} m_0 + t] \end{aligned} \quad (7)$$

이다. 단, $R=R_1$, $t=-Rt_0+t_1$ 이다.

위에서 구한 optical flow $[u, v] = [x_1-x_0, y_1-y_0]$ 와 식 (7)를 이용해서 다음과 같이 쓸 수 있다.

$$\begin{aligned} u &= \frac{x_0 r_{11} + y_0 r_{12} + f r_{13} + \frac{f}{s} t_x}{\frac{x_0}{f} r_{31} + y_0 r_{32} + f r_{33} + \frac{1}{s} t_z} - x_0 \\ v &= \frac{x_0 r_{21} + y_0 r_{22} + f r_{23} + \frac{f}{s} t_y}{\frac{x_0}{f} r_{31} + y_0 r_{32} + f r_{33} + \frac{1}{s} t_z} - y_0 \end{aligned} \quad (8)$$

여기서, 각각 u 와 v 는 s (depth)의 함수이다. 하지만, 주어진 조건으로부터 s 를 알 수 없으므로 식(8)에서 s 를 소거하고 정리하면 다음과 같은 비선형 방정식을 얻을 수 있다.

$$\begin{aligned} & t_x(fx_0 r_{21} + fy_0 r_{22} + f^2 r_{23} - fy_0 r_{33} - fur_{33} \\ & - x_0 y_0 r_{31} - y_0^2 r_{32} - ux_0 r_{31} - vy_0 r_{32}) \\ & - t_y(fx_0 r_{11} + fy_0 r_{12} + f^2 r_{13} - fx_0 r_{33} - fur_{33} \\ & - x_0^2 r_{31} - x_0 y_0 r_{32} - ux_0 r_{31} - uy_0 r_{32}) \\ & + t_z(ux_0 r_{11} + vy_0 r_{12} + fur_{13} + x_0 y_0 r_{11} + y_0^2 r_{12} \\ & + fy_0 - ux_0 r_{21} - uy_0 r_{22} - fur_{23} - x_0^2 r_{21} \\ & - x_0 y_0 r_{22} - fx_0 r_{23}) = 0 \end{aligned} \quad (9)$$

여기서 회전 행렬 R 은 3차원에 대한 회전으로 각각 좌표계 축 e_1, e_2, e_3 에 대한 세계의 연속적인 회전으로 각 α, β, γ 로 표현할 수 있다. 회전 행렬 R 은 3개의 미지수로 이루어지며 e_1 를 중심으로 회전하는 회전 행렬 R_1 로 나타낼 수 있다.

$$\begin{aligned} R_1(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \\ R_2(\beta) &= \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \\ R_3(\gamma) &= \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (10)$$

결국, 우리가 구하고자 하는 회전 행렬은 식 (11)과 같은 세 개의 미지수로 나타낼 수 있다[7].

$$R = R_1 R_2 R_3 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$\begin{aligned} \text{where } r_{11} &= \cos \beta \cos \gamma \\ r_{12} &= -\cos \beta \sin \gamma \\ r_{13} &= \sin \beta \\ r_{21} &= \sin \alpha \sin \beta \cos \gamma + \cos \alpha \sin \gamma \\ r_{22} &= -\sin \alpha \sin \beta \sin \gamma + \cos \alpha \cos \gamma \\ r_{23} &= -\sin \alpha \cos \beta \\ r_{31} &= -\cos \alpha \sin \beta \cos \gamma + \sin \alpha \sin \gamma \\ r_{32} &= \cos \alpha \sin \beta \sin \gamma + \sin \alpha \cos \gamma \\ r_{33} &= \sin \alpha \cos \beta \end{aligned} \quad (11)$$

식(11)를 식 (9)에 대입하면 우리는

$f(\alpha, \beta, \gamma, t_x, t_y, t_z)=0$ 이라는 6개의 미지수를 가진 비선형 방정식을 얻을 수 있다.

IV. 3차원 카메라 운동 추정

일반적인 비선형 방정식을 푸는 방법으로는 Newton-Raphson나 Newton's Rule[8]이 있는데, 이 방법으로 $f(\alpha, \beta, \gamma, t_x, t_y, t_z) = 0$ 을 풀면 trivial solution, $t_x = t_y = t_z = 0$ 을 얻는다. trivial solution이 생기는 이유는 식(8)에서 depth 정보를 소거하면서 분수식 $\frac{b}{a} = \frac{d}{c}$ 를 $ad=bc$ 로 정리했기 때문이다.

$\frac{b}{a} = \frac{d}{c}$ 는 $ad=bc$ 의 필요충분조건이 아니라 충분조건이므로 이 문제를 해결하기 위해서 조건 $t_x^2 + t_y^2 + t_z^2 > 0$ 과 $0 < \alpha < 2\pi, 0 < \beta < 2\pi, 0 < \gamma < 2\pi$ 을 추가하였다.

이러한 방정식은 운동 정보의 추정 문제는 파라미터 $\alpha, \beta, \gamma, t_x, t_y, t_z$ 로 주어진 조건에서 함수 f 를 최소화하는 문제로 바뀔 수 있다.

$$\begin{aligned} \min_{\alpha, \beta, \gamma, t_x, t_y, t_z} & |f(\alpha, \beta, \gamma, t_x, t_y, t_z)| \\ \text{subject to } & t_x^2 + t_y^2 + t_z^2 > 0 \end{aligned} \quad (12)$$

본 논문에서는 rate-constrained problem을 unconstrained problem으로 바꾸어 주는 효과적인 방법으로 알려져 있는 Lagrange multiplier를 이용하여 주어진 문제를 다음과 같이 수식화 하였다.

$$\begin{aligned} \min_{\alpha, \beta, \gamma, t_x, t_y, t_z} & [E(\lambda) = |f(\alpha, \beta, \gamma, t_x, t_y, t_z)| \\ & + \lambda e^{-\alpha(t_x^2 + t_y^2 + t_z^2)}] \end{aligned} \quad (13)$$

단, 여기서 λ 는 Lagrange multiplier이고,

$e^{-\alpha(t_x^2 + t_y^2 + t_z^2)}$ 는 Gaussian 함수이다.

식 (13)을 풀기 위해서 본 논문에서는 iterative method를 사용하였다. 추가 조건으로부터 trivial solution을 제거하고 조건이 있는 최소화 문제를 Lagrange multiplier를 통해서 적절한 해를 다음과 같은 과정으로 찾아내었다.

1. 오차를 ΔE 로 놓고 $\alpha, \beta, \gamma, t_x, t_y, t_z$ 의 초기 값을 지정한다.
2. 영상의 각 화소점에서 구한 optical flow u, v 를 이용하여 $\alpha, \beta, \gamma, t_x, t_y, t_z$ 를 Newton's Rule을 이용하여 구한다.
3. 위에서 구한 $\alpha, \beta, \gamma, t_x, t_y, t_z$ 를 이용해서 $E(\lambda)$ 를 계산한다.
4. 전체적인 오차 범위 안에 들 때까지 1에서 3까지의 과정을 반복한다.
5. $E(\lambda)$ 를 최소화하는 $\alpha, \beta, \gamma, t_x, t_y, t_z$ 를 구한다.

V. 실험 결과

Lagrange multiplier를 이용한 운동 정보의 추정
은 합성 영상으로 실험되었다. 합성 영상은 임의로
병진과 회전 운동을 한 카메라를 이용해서 연속적으
로 여러 장을 취득하였다.

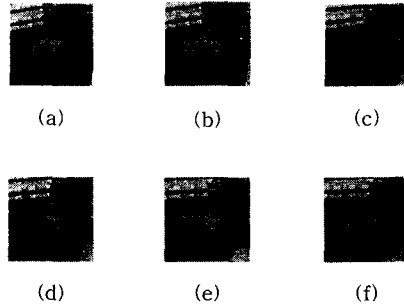


그림 2. 임의의 카메라 운동으로 얻은 연속적
인 합성 영상.

대부분의 iterative method는 초기값에 따라서 그
결과가 많은 차이가 생긴다. 이렇게 생기는 오차를
최소화하기 위해서 우선 첫 번째 영상에서 주어진
모델의 특징점과 optical flow를 통해서 각 좌표를
normalize한다. 이렇게 좌표계를 normalize하면 t_x , t_y
는 각각 normalize된 회전 인자 \tilde{r}_{13} , \tilde{r}_{23} 의 함수로
바뀔 수 있다. t_x , t_y 가 회전 인자의 함수라는 조건을
추가시키고, 위에서 제안한 3차원 카메라 운동 추정
방법을 사용하여 표 2와 같은 결과를 얻었다.

	Rotation axis			Translation		
	α	β	γ	t_x	t_y	t_z
(a), (b)	0.0154	0.0492	0.0359	-2.5094	0.7848	-1.0425
(b), (c)	0.0140	0.0519	0.0391	-2.5889	0.6987	-0.8996
(c), (d)	0.0122	0.0544	0.0423	-2.6612	0.5977	-0.7487
(d), (e)	0.0101	0.0567	0.0453	-2.7236	0.4825	-0.5905
(e), (f)	0.0075	0.0585	0.0479	-2.7735	0.3547	-0.4264

표 1. 연속적인 합성영상에서의 회전과 병진 인자
의 이론치.

	Rotation axis			Translation		
	α	β	γ	t_x	t_y	t_z
(a), (b)	0.0350	0.0965	0.0429	-2.5094	0.7830	-1.0574
(b), (c)	0.0341	0.1077	0.0347	-2.5872	0.6891	-1.0136
(c), (d)	0.0323	0.1205	0.0253	-2.6568	0.5913	-1.0038
(d), (e)	0.0297	0.1373	0.0148	-2.7193	0.4794	-0.9465
(e), (f)	0.0264	0.1634	0.0035	-2.7721	0.3426	-0.8249

표 2. 논문에서 제안한 방법으로 구한 회전과 병진
인자의 실험치.

IV. 결 론

본 논문은 Lagrange multiplier를 사용해서 카메라
의 3차원 운동을 추정하는 새로운 방법을 제안하였
다. 이 방법은 3차원 모델과 배경의 많은 특징들을
이용하는 optical flow를 구한 후, 이로부터 얻어진
속도 정보와, 카메라와 3차원 물체로부터 비선형 방
정식을 모델링하여 병진과 회전인자를 추정하는 방
법이다. 본 논문에서는 회전 인자를 선형화하지 않았
기 때문에 카메라의 회전 인자가 큰 경우에도 정확
한 결과를 얻을 수 있었다.

참고문헌

- [1] H. C. Longuet-Higgins, "A Computer Program for reconstructing a scene from two projections", *Nature*, Vol. 293, p. 133-135, 1981
- [2] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-d motion parameters of rigid bodies with curved surfaces", *IEEE Trans. Pattern Anal., and Machine Intel.*, Vol. PAMI-6(1), p. 13-27, 1984
- [3] O. Faugeras, *Three-dimensional computer vision : a geometric view point*, MIT Press, 1993
- [4] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images-A Review" *Proceedings of the IEEE*, Vol. 76, NO. 8, p. 917-935, August, 1988
- [5] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "Performance of optical flow techniques", *International Journal of Computer Vision*, Vol. 12:1, p. 43-77, 1994
- [6] B. D. Lucas and T. Kanade "An iterative image registration technique with an application to stereo vision", *Proc. DARPA IU Workshop*, p. 121-131, 1981
- [7] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision", Prentice Hall, 1998
- [8] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in C", Cambridge University Press, 1988