LETTER

Moving Object Detection from Optical Flow without Empirical Thresholds *

Naoya OHTA†, **Kenichi KANATANI**†, *Members*, *and* **Kazuhiro KIMURA**††, *Nonmember*

SUMMARY We show that moving objects can be detected from optical flow without using any knowledge about the magnitude of the noise in the flow or any thresholds to be adjusted empirically. The underlying principle is viewing a particular interpretation about the flow as a *geometric model* and comparing the relative "goodness" of candidate models measured by the *geometric AIC*.

key words: geometric AIC, model selection, optical flow, moving object detection, statistical inference

1. Introduction

In autonomous vehicle navigation by visual sensing, detecting objects that are moving independently of the background, such as people and other vehicles, is of utmost importance. This type of "intelligent processing" has been usually done by incorporating various thresholds to be adjusted empirically [3], [5], [8]. They depend on the environment, the device, and the image processing techniques by which the images are obtained. In particular, they heavily depend on the magnitude of noise: the decision should be strict for a low noise level; a large deviation should be tolerated for a high noise level. Hence, values adjusted in one environment become meaningless in another environment.

In this letter, we show that moving objects can be detected from optical flow without using any knowledge about the magnitude of the noise in the flow or any thresholds to be adjusted empirically. The underlying principle is viewing a particular interpretation about the flow as a *geometric model* and comparing the relative "goodness" of candidate models measured by the *geometric AIC* [2] obtained by modifying the *AIC* (*Akaike information criterion*) [1].

2. Models for Optical Flow

By a *model* of optical flow, we mean "the condition that would be satisfied if noise were not present". If we define an XYZ coordinate system such that the origin is at the center of the lens and the Z-axis is in the direction of the optical axis and adopt the focal length as the unit of length, we can identify $Z = 1$ as the image plane. Suppose we observe flow $(\dot{x}, \dot{y})$ at point (x, y). We represent the flow and the position by the following three-dimensional vectors:

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad \dot{\boldsymbol{x}} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ 0 \end{pmatrix}. \tag{1}$$

Let $\boldsymbol{v}$ and $\boldsymbol{\omega}$ be, respectively, the instantaneous translation velocity and rotation velocity of the camera; we call $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ the *motion parameters*. Suppose we observe flow $\{\dot{\boldsymbol{x}}_\alpha\}$ at points $\{\boldsymbol{x}_\alpha\}$, $\alpha = 1, ..., N$, and let $\{\bar{\dot{\boldsymbol{x}}}_\alpha\}$ be the (unknown) true flow. For a stationary scene, consider the following models:

1. **General motion model** $\mathcal{S}_{\rm gm}$: The motion parameters $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ are unconstrained. In the absence of noise, we have the following *epipolar equation* [2], [6] ($|\cdot, \cdot, \cdot|$ denotes scalar triple product):

$$|\boldsymbol{x}_\alpha, \bar{\dot{\boldsymbol{x}}}_\alpha, \boldsymbol{v}| + (\boldsymbol{v} \times \boldsymbol{x}_\alpha, \boldsymbol{\omega} \times \boldsymbol{x}_\alpha) = 0. \tag{2}$$

2. **General translation model** $\mathcal{S}_{\rm gt}$: The camera translates without rotation. In the absence of noise, we have

$$|\boldsymbol{x}_\alpha, \bar{\dot{\boldsymbol{x}}}_\alpha, \boldsymbol{v}| = 0. \tag{3}$$

3. **Special translation model** $\mathcal{S}_{\rm st}$: The camera undergoes a known translation $\boldsymbol{v}^0$ without rotation. In the absence of noise, we have

$$|\boldsymbol{x}_\alpha, \bar{\dot{\boldsymbol{x}}}_\alpha, \boldsymbol{v}^0| = 0. \tag{4}$$

3. Geometric AIC and Model Comparison

Let $\epsilon^2 V_0[\dot{\boldsymbol{x}}_\alpha]$ be the covariance matrix of flow $\dot{\boldsymbol{x}}_\alpha$. We call $V_0[\dot{\boldsymbol{x}}_\alpha]$ the *normalized covariance matrix*, and ϵ, which represents unknown noise magnitude, the *noise level*. The normalized covariance matrix can be easily determined if the optical flow is detected by applying the so-called gradient constraint [4], [7]. The goodness of a model is measured by the *geometric AIC*[2].

1. **General motion model** $\mathcal{S}_{\rm gm}$: An optimal estimate of $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ is obtained by minimizing the following function [2], [6]:

$$J_{\rm gm}[\boldsymbol{v}, \boldsymbol{\omega}] = \sum_{\alpha=1}^{N} \frac{(|\boldsymbol{x}_\alpha, \dot{\boldsymbol{x}}_\alpha, \boldsymbol{v}| + (\boldsymbol{v} \times \boldsymbol{x}_\alpha, \boldsymbol{\omega} \times \boldsymbol{x}_\alpha))^2}{(\boldsymbol{v}, (\boldsymbol{x}_\alpha \times V_0[\dot{\boldsymbol{x}}_\alpha] \times \boldsymbol{x}_\alpha)\boldsymbol{v})}. \tag{5}$$

Manuscript received June 4, 1997.
Manuscript revised September 5, 1997.

†The authors are with the Department of Computer Science, Gunma University, Kiryu-shi, 376 Japan.
††The author is with CSK, Co., Tokyo, 163-02 Japan.

*This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 09680352).

Let $\hat{J}_{\rm gm}$ be its residual. Since the constraint (2) has rank 1 and five degrees of freedom, the geometric AIC is

$$AIC(\mathcal{S}_{\rm gm}) = \hat{J}_{\rm gm} + 2(N+5)\epsilon^2. \tag{6}$$

An unbiased estimator of ϵ^2 is obtained as follows:

$$\hat{\epsilon}^2 = \frac{\hat{J}_{\rm gm}}{N-5}. \tag{7}$$

2. **General translation model** $\mathcal{S}_{\rm gt}$: An optimal estimate of $\boldsymbol{v}$ is obtained by minimizing

$$J_{\rm gt}[\boldsymbol{v}] = \sum_{\alpha=1}^{N} \frac{|\boldsymbol{x}_\alpha, \dot{\boldsymbol{x}}_\alpha, \boldsymbol{v}|^2}{(\boldsymbol{v}, (\boldsymbol{x}_\alpha \times V_0[\dot{\boldsymbol{x}}_\alpha] \times \boldsymbol{x}_\alpha)\boldsymbol{v})}. \tag{8}$$

Let $\hat{J}_{\rm gt}$ be its residual. Since the constraint (3) has rank 1 and two degrees of freedom, the geometric AIC is

$$AIC(\mathcal{S}_{\rm gt}) = \hat{J}_{\rm gt} + 2(N+2)\epsilon^2. \tag{9}$$

An unbiased estimator of ϵ^2 is obtained as follows:

$$\hat{\epsilon}^2 = \frac{\hat{J}_{\rm gt}}{N-2}. \tag{10}$$

3. **Special translation model** $\mathcal{S}_{\rm st}$: Let

$$\hat{J}_{st} = \sum_{\alpha=1}^{N} \frac{|\boldsymbol{x}_\alpha, \dot{\boldsymbol{x}}_\alpha, \boldsymbol{v}^0|^2}{(\boldsymbol{v}^0, (\boldsymbol{x}_\alpha \times V_0[\dot{\boldsymbol{x}}_\alpha] \times \boldsymbol{x}_\alpha)\boldsymbol{v}^0)}. \tag{11}$$

The constraint (4) has rank 1, but no unknowns are involved. So, the geometric AIC is

$$AIC(\mathcal{S}_{\rm st}) = \hat{J}_{\rm st} + 2N\epsilon^2. \tag{12}$$

1. $\mathcal{S}_{\rm gm}$ vs. $\mathcal{S}_{\rm gt}$: The square noise level is estimated from Eq. (7). The general translation model $\mathcal{S}_{\rm gt}$ is preferred if

$$K_{\rm gt} = \sqrt{\frac{N-5}{3N+5}\left(\frac{\hat{J}_{\rm gt}}{\hat{J}_{\rm gm}} + \frac{2(N+2)}{N-5}\right)} < 1. \tag{13}$$

2. $\mathcal{S}_{\rm gt}$ vs. $\mathcal{S}_{\rm st}$: The square noise level is estimated from Eq. (10). The special translation model $\mathcal{S}_{\rm st}$ is preferred if

$$K_{\rm st} = \sqrt{\frac{N-2}{3N+2}\left(\frac{\hat{J}_{\rm st}}{\hat{J}_{\rm gt}} + \frac{2N}{N-2}\right)} < 1. \tag{14}$$

4. Moving Object Detection

Suppose a camera is fixed to a vehicle moving ahead. The camera observes a special translation flow if the scene is stationary. If an object is translating without rotating in the scene, the camera observes a general translation flow in the object region. Hence, we can detect a moving object by sliding a window over the image frame and comparing the special translation model $\mathcal{S}_{\rm st}$ with the general translation model $\mathcal{S}_{\rm gt}$ within the window. We judge that no object is moving if $\mathcal{S}_{\rm st}$ is favored. In other words, we compute Eq. (14) and test if $K_{\rm st} < 1$.

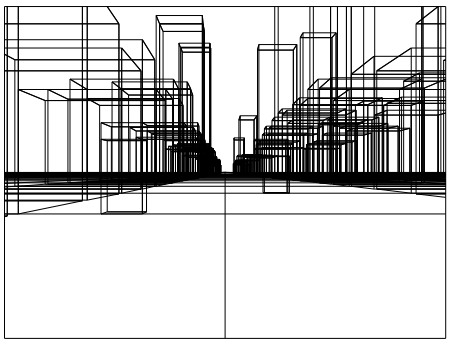

Fig. 1 A wireframe model of a scene.

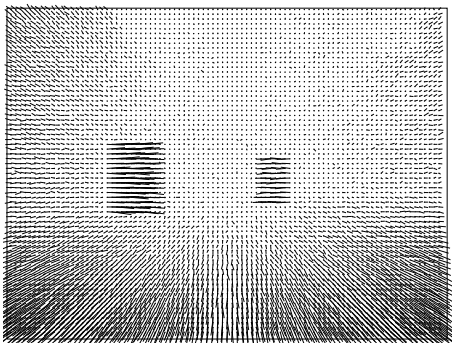

Fig. 2 Synthetic optical flow.

Admitting the fact that numerous sources of uncertainty that are not modeled in the theory exist, it is more realistic to regard the value $K_{\rm st}$ as the "degree of non-existence of moving objects".

Figure 1 is a simulated frontal view of a wire-frame urbane scene observed from a vehicle running ahead with 40 km/h. The image size is 640 × 480 with focal length 600 pixels and view angle 56.1° × 43.6°. Rectangular objects are moving with 4 km/h orthogonal to the vehicle motion in the stationary background. Figure 2 shows optical flow generated by adding Gaussian noise of mean 0 and standard deviation 0.2 (pixels) to the theoretical flow defined in every seven pixels (the flow is magnified 1.5 times in the figure).

Figure 3 is a gray-level image of $K_{\rm st}$ computed at the center of a 5×5 window sliding over the image (black for $K_{\rm st} = 0.95$ and white for $K_{\rm st} = 2$), and Fig. 4 is a binary image of $K_{\rm st} > 1$. The value of $K_{\rm st}$ was computed by a technique called *renormalization* [6] with a modification that flow components that give the largest values of a fixed percentage in eq. (8) are removed in the computation.

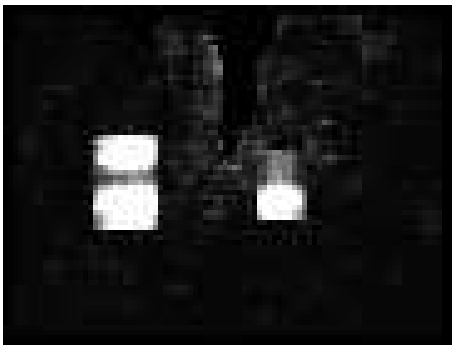

Fig. 3 Gray-level image of $K_{\rm st}$.

Fig. 4 Binary image of $K_{\rm st} > 1$.

5. Concluding Remarks

The criterion for moving object detection described in this letter has the following features:

1. It does not involve any thresholds to be adjusted empirically.

2. It does not require any prior knowledge of the magnitude of noise.

We can see from Figs. 3 and 4 that a threshold slightly larger than 1 would detect moving objects more distinctly. Perhaps this is because our theory involves linear approximation based on the assumption that the noise is small [2]. We also applied our inference to a small window size (25 pixels) in our experiment. In general, statistical inference based on information criteria may not necessarily be reliable and the solution may be somewhat biased when the noise is not small. Hence, we need some kind of empirical adjustments in real applications. The performance will also be improved if information about the environment is incorporated. We have shown, however, that at least in the most fundamental level the judgment can be done without empirical thresholds or knowledge about the noise level.

References

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automation Control* vol.19, no.6, pp.176–723, 1974.

[2] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, 1996.

[3] A. Nagai, Y. Kuno and Y. Shirai, "Detection of moving objects in changing background" (in Japanese), *Trans. IEICE*, vol.J80-D-II, no.5, pp.1986–1095, 1997.

[4] N. Ohta, "Image movement detection with reliability indices," *IEICE Trans. Inf. & Syst.*, vol.E74-D, no.10, pp.3379–2288, 1991.

[5] N. Ohta, "Structure from motion with confidence measure and its application for moving object detection" (in Japanese), *IEICE Trans.*, vol.76-D-II, no.8, pp. 1562–1571, 1993.

[6] N. Ohta and K. Kanatani, "Optimal structure from motion algorithm for optical flow," *IEICE Trans. Inf. & Syst.*, vol.E78-D, no.12, pp.1559–1566, 1995.

[7] N. Ohta, "Optical flow detection using general noise model," *IEICE Trans. Inf. & Syst.*, vol.E79-D, no.7, pp.951–957, 1996.

[8] M. Watanabe, N. Takeda and K. Onoguchi, "A moving object recognition method by the principal component" (in Japanese), IPSJ Sig Notes 96-CV-99, March 1996.