Robust Digital Image Stabilization Using the Kalman Filter*

ISL Lab Seminar
Han-Sol Kang

* Wang, Chuntao, et al. "Robust digital image stabilization using the Kalman filter." Consumer Electronics, IEEE Transactions on 55.1 (2009): 6-14.

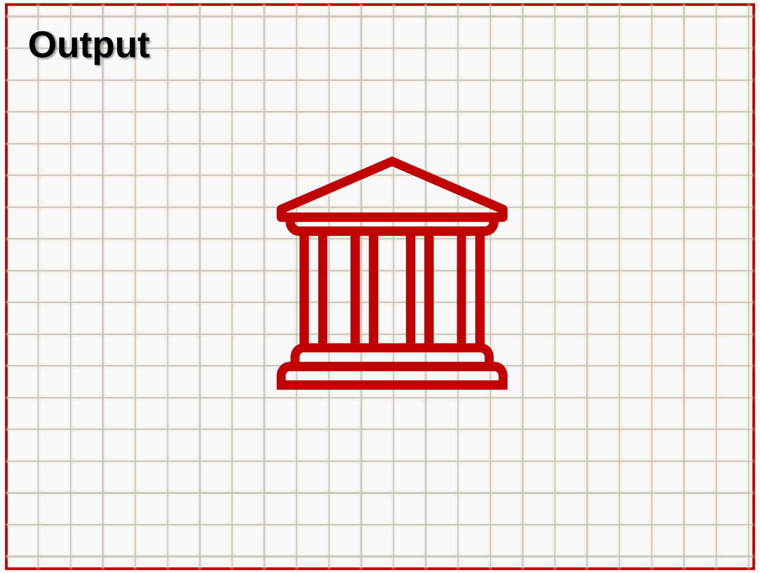
INDEX

01 Introduction

Image Stabilization

[Hand held]

[Mounted on vehicle]



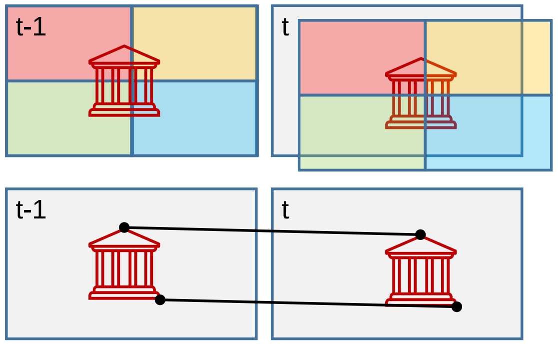
01 Introduction

☆ Image Stabilization

01 Introduction

☆ Type of Image Stabilization

- Image Stabilization (IS)
  - Optical IS (OIS)
  - Electric IS (EIS)
  - Digital IS (DIS)
    - Block Matching Algorithm (BMA)
    - Feature Based Matching (FBM)

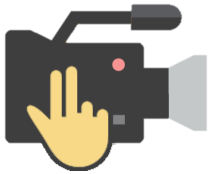


02 Camera Trajectory Model

☆ Camera Trajectory Model

Camera Trajectory : Represents the intentional motion.

Camera shake : Noise version of intentional motion.



⇠- - -⇢ : Camera shake

[Dynamic equation]

$$X_t = AX_{t-1} + W_t$$
$$Z_t = HX_t + V_t$$

$X_t (t = 0, 1, \cdots)$: the state at time t that represents the camera position, velocity, and acceleration

$Z_t (t = 0, 1, \cdots)$: the observed camera position at time t

A, H : the state transition matrix and the measurement one, respectively

Assume that $W_t \sim N(0, Q_t)$ and $V_t \sim N(0, R_t)$ are white Gaussian noise that are independent of each other

X : Intentional motion
Z : Global motion

02 Camera Trajectory Model

Camera Trajectory Model

[Initial expected value]

$$\hat{X}_0^+ = E(X_0)$$

$$P_0^+ = E\left[(X_0 - \hat{X}_0^+)(X_0 - \hat{X}_0^+)^T\right]$$

- : *prior*
+ : *posterior*

[Prediction step]

$$\hat{X}_t^- = A\hat{X}_{t-1}^+$$

$$P_t^- = AP_{t-1}^+A^T + Q_{t-1}$$

[Correction step]

$$K_t = P_t^- H^T [HP_t^- H^T + R_t]^{-1}$$

$$P_t^+ = [I - K_t H]P_t^-$$

$$\hat{X}_t^+ = \hat{X}_t^- + K_t[Z_t - H\hat{X}_t^-]$$

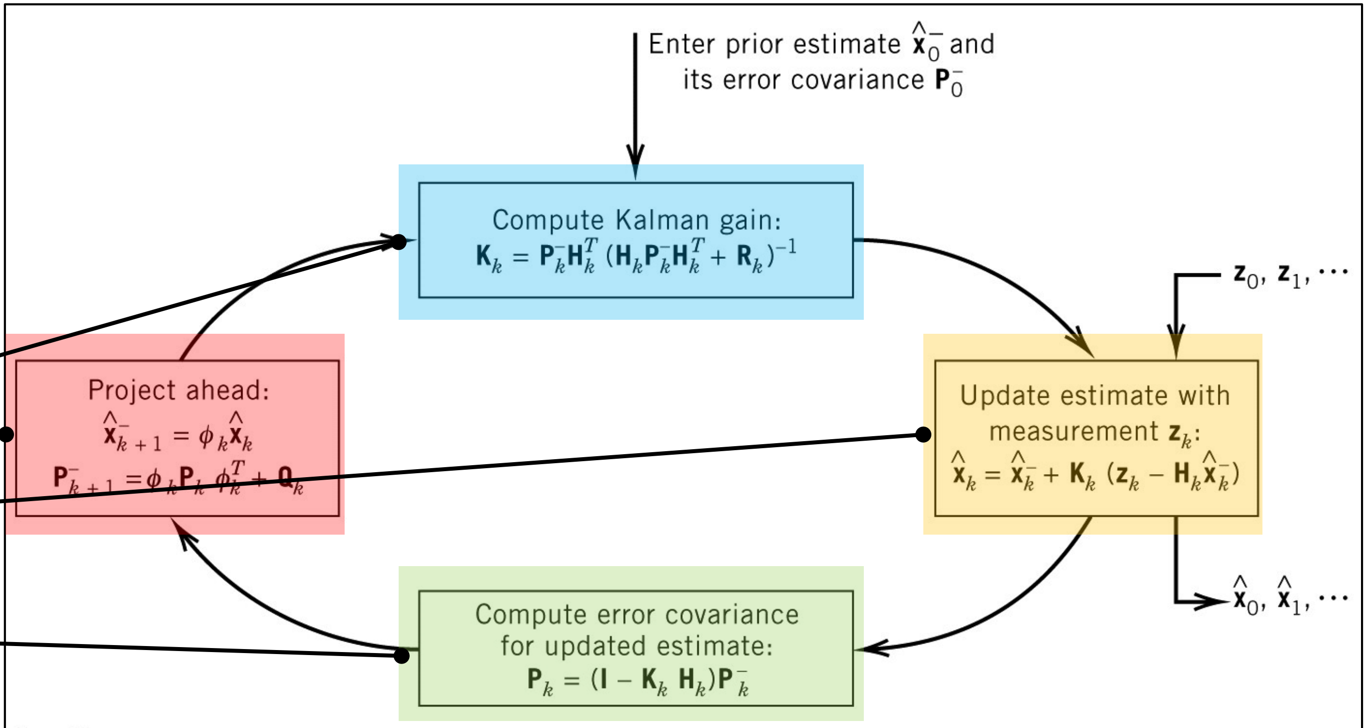


Figure 4.1

02 Camera Trajectory Model

Motion Model

[Affine transform]

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- Rotation
- Translation
- Scaling
- Shearing

Rigid Body Transformation : Translation + Rotation

Similarity Transformation : Translation + Rotation + Uniform scaling

Affine Transformation : Translation + Rotation + Scaling + Shearing

[Rigid Body Transformation (2-D)]

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} T_t^x \\ T_t^y \end{bmatrix}$$

(x_t, y_t) : the coordinate at time t

θ_t : the rotation angle

T_t^x, T_t^y : the translation in the horizontal and vertical direction, respectively

$$\Rightarrow U_t = F_t U_{t-1} + T_t$$

[Accumulative motion model]

$$U_t = \left(\prod_{k=1}^{t} F_k\right) U_0 + \sum_{k=1}^{t}\left(\prod_{j=k+1}^{t} F_j\right) T_k = F_t^A U_0 + T_t^A$$

where $\prod_{j=k+1}^{t} F_j = \begin{pmatrix} \cos\theta' & -\sin\theta' \\ \sin\theta' & \cos\theta' \end{pmatrix}$ $\quad \theta' = \sum_{j=k+1}^{t} \theta_j$

02 Camera Trajectory Model

Design of the KF

[Rotation angle at time t]

$$\hat{\theta}_t = \hat{\theta}_{t-1} + n_t$$

$n_t \sim N(0, \sigma_r)$: white noise with variance σ_r

$$X_t = AX_{t-1} + W_t \Rightarrow \begin{bmatrix} \hat{\theta}_t \\ \hat{T}_t^x \\ \hat{T}_t^{xv} \\ \hat{T}_t^y \\ \hat{T}_t^{yv} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\theta}_{t-1} \\ \hat{T}_{t-1}^x \\ \hat{T}_{t-1}^{xv} \\ \hat{T}_{t-1}^y \\ \hat{T}_{t-1}^{yv} \end{bmatrix} + \begin{bmatrix} n_t^\theta \\ 0 \\ n_t^v \\ 0 \\ n_t^v \end{bmatrix}$$

[Translation (assume constant velocity)]

$$\hat{T}_t^v = \hat{T}_{t-1}^v + n_t^v \qquad \hat{T}_t^v = \begin{bmatrix} \hat{T}_t^{xv} \\ \hat{T}_t^{yv} \end{bmatrix}$$

$$\hat{T}_t = \hat{T}_{t-1} + \hat{T}_t^v$$

$n_t^v \sim N(0, \sigma_T^v)$: white noise with variance σ_T^v

$$Z_t = HX_t + V_t \Rightarrow \begin{bmatrix} \theta_t^o \\ T_t^{ox} \\ T_t^{oy} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} X_t + \begin{bmatrix} n_t^{o\theta} \\ n_t^{oT} \\ n_t^{oT} \end{bmatrix}$$

The auxiliary variable $\hat{T}_t^v$ cannot be observed in Z_t

Superscript "o" denotes the variables for observed

$n_t^{o\theta} \sim N(0, \sigma_r^o)$ and $n_t^{oT} \sim N(0, \sigma_T^o)$ are white Gaussian noise

02 Camera Trajectory Model

Design of the KF

$$Q_t = diag(\sigma_r, 0, \sigma_T^v, 0, \sigma_T^v)$$

$$R_t = diag(\sigma_r^o, \sigma_T^o, \sigma_T^o)$$

$$\hat{X}_0^+ = (0,0,0,0,0)^T$$

$$P_0^+ = diag(\sigma_r, \sigma_T, \sigma_T^v, \sigma_T, \sigma_T^v)$$

$$\begin{bmatrix} \hat{\theta}_t \\ \hat{T}_t^x \\ \hat{T}_t^{xv} \\ \hat{T}_t^y \\ \hat{T}_t^{yv} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\theta}_{t-1} \\ \hat{T}_{t-1}^x \\ \hat{T}_{t-1}^{xv} \\ \hat{T}_{t-1}^y \\ \hat{T}_{t-1}^{yv} \end{bmatrix} + \begin{bmatrix} n_t^\theta \\ 0 \\ n_t^v \\ 0 \\ n_t^v \end{bmatrix}$$

$$\begin{bmatrix} \theta_t^o \\ T_t^{ox} \\ T_t^{oy} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} X_t + \begin{bmatrix} n_t^{o\theta} \\ n_t^{oT} \\ n_t^{oT} \end{bmatrix}$$

$$\hat{X}_0^+ = E(X_0)$$

$$P_0^+ = E\left[(X_0 - \hat{X}_0^+)(X_0 - \hat{X}_0^+)^T\right]$$

$diag(\ )$: diagonal matrix

[Z_t accumulative motion]

$$\begin{bmatrix} \theta_t^o \\ T_t^{ox} \\ T_t^{oy} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} X_t + \begin{bmatrix} n_t^{o\theta} \\ n_t^{oT} \\ n_t^{oT} \end{bmatrix}$$

$$\theta_t^o = \sum_{j=1}^{t} \theta_j = \theta_{t-1}^o + \tilde{\theta}_t$$

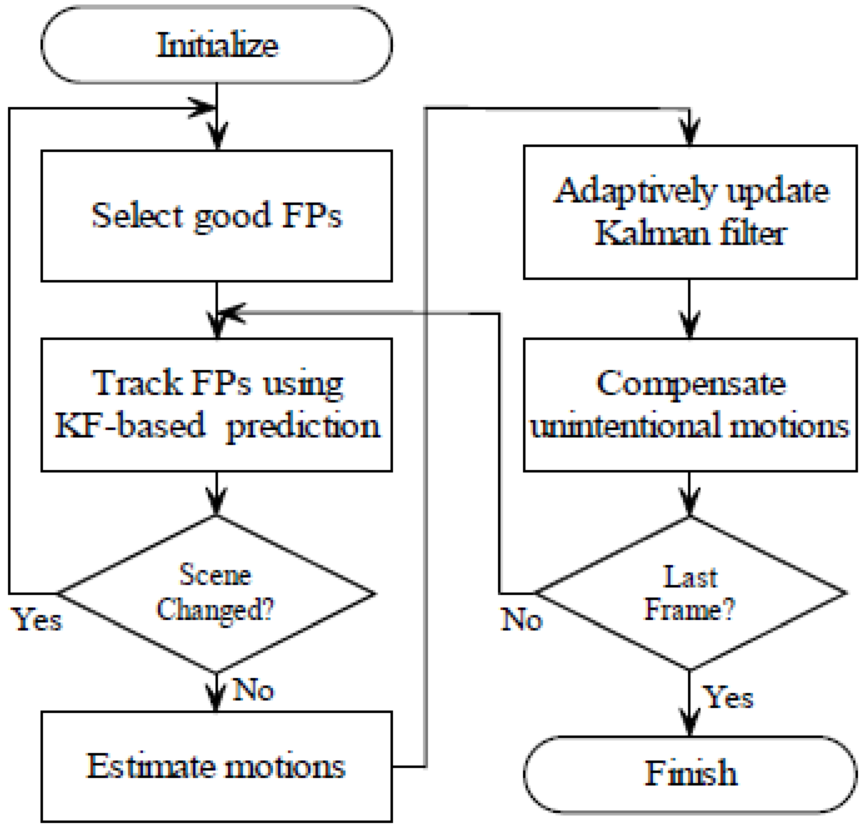
$$T_t^o = \sum_{k=1}^{t} \left(\prod_{j=k+1}^{t} F_j \right) T_k = F_t T_{t-1}^o + \tilde{T}_t$$

where $\prod_{j=k+1}^{t} F_j = \begin{pmatrix} \cos\theta' & -\sin\theta' \\ \sin\theta' & \cos\theta' \end{pmatrix}$ $\quad \theta' = \sum_{j=k+1}^{t} \theta_j$

$\tilde{\theta}_t$ and $\tilde{T}_t$ are computed using $\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} T_t^x \\ T_t^y \end{bmatrix}$

03 FB-Based DIS Algorithm

☆ The process of the proposed DIS algorithm



03 FB-Based DIS Algorithm

Selection of Good FPs

$$G = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} I_x^2(x,y) & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y^2(x,y) \end{bmatrix}$$

(p_x, p_y) : the point (x, y)

w_x, w_y : the searching window size

$I_x(x,y),\ I_y(x,y)$: spatial derivative of the image frame in the horizontal (vertical) direction

$$\min(\lambda_1, \lambda_2) > \lambda_{thresh}$$

$\lambda_1,\ \lambda_2$: eigen-values of G

λ_{thresh} : generally set as $\lambda_{thresh} = r\lambda_{\max} (0 < r < 1)$ $\lambda_{\max}$: maximum eigen-value obtained from eigen-values of all points

03 FB-Based DIS Algorithm

FP Tracking Using KF-Based Prediction

[Current camera motion $\tilde{\mathrm{M}}_t$]

$$\theta_t^o = \theta_{t-1}^o + \tilde{\theta}_t$$
$$T_t^o = F_t T_{t-1}^o + \tilde{T}_t$$

$\tilde{\mathrm{M}}_t = [\tilde{\theta}_t, \tilde{\mathrm{T}}_t]^T$ is obtained by solving the equation $\mathrm{M}_t^o = \mathrm{M}_{t-1}^o + \tilde{\mathrm{M}}_t$

where the related parameters $\mathrm{M}_t^o = [\theta_t^o, T_t^o]^T$ and $\mathrm{M}_{t-1}^o = [\theta_{t-1}^o, T_{t-1}^o]^T$ are extracted from $\hat{X}_t^-$ and $\hat{X}_{t-1}^+$, respectively

[FP $(x_t^i,\ y_t^i)$]

$(x_t^i,\ y_t^i)$ is predicted by performing $\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} T_t^x \\ T_t^y \end{bmatrix}$

where $(x_t^i,\ y_t^i)\ (1 \le i \le N_0)$

Finally, the **predicted FP** instead of the previous one $(x_{t-1}^i,\ y_{t-1}^i)$ is set as the initial position for the correspondence tracking, which reduces the searching range and thus **speeds up the tracking process**

03 FB-Based DIS Algorithm

Detection of Scene Change

Assume that N_t pairs of FPs are obtained after finding the correspondence.

$N_t \geq N_{thresh}$: no scene change

If scene change is detected,

Current image frame ➡ new reference frame
FPs : re-selection
Parameter of KF : re-initialization

[Irregular condition and the lack of features]

$N_t < N_{thresh}$: scene change?? X

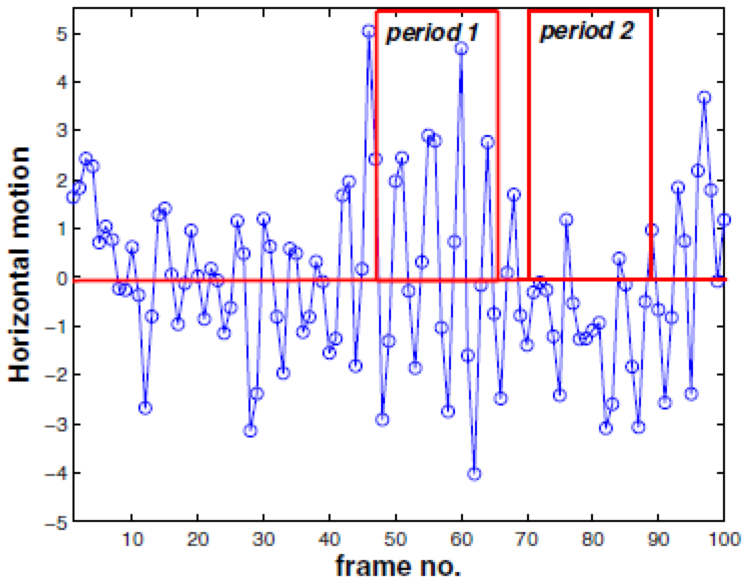
$\Delta t \geq \Delta T_{thresh}$: reference frame change Δt: time difference between the current detection of scene change and the previous one

03 FB-Based DIS Algorithm

☆ Adaptive Measurement-Update of KF

$R_t = diag(\sigma_r^o, \sigma_T^o, \sigma_T^o)$ x ➡ the practical camera shake at different time may be greatly different
We develop a new scheme, called and **adaptive KF**

the camera shake results in the sign change of the estimated motion $\tilde{M}_t$ offset by their mean values



zero-crossing cases : period 1 > period 2

➡ camera shakes : period 1 > period 2

$$R_t = \begin{cases} R_{t4} & if \ \ C_t > CT_4 \\ R_{t3} & if \ \ CT_3 < C_t \le CT_4 \\ R_{t2} & if \ \ CT_2 < C_t \le CT_3 \\ R_{t1} & if \ \ CT_1 < C_t \le CT_2 \\ R_{t0} & \text{otherwise} \end{cases}$$

C_t : the number of zero-crossing

$R_{ti} (i = 0 \cdots 4)$: measurement noise with different variance

$CT_i (i = 1 \cdots 4)$: predefined threshold

03 FB-Based DIS Algorithm

Compensation of Unintentional Motion

The global motion M_t^o contains the intentional motion $\hat{\mathrm{M}}_\mathrm{t}$ and the unintentional one M_t^c

M_t^c should be computed first.

accumulative motion model $U_t = \left(\prod_{k=1}^{t} F_k\right) U_0 + \sum_{k=1}^{t}\left(\prod_{j=k+1}^{t} F_j\right) T_k = F_t^A U_0 + T_t^A$

$$U_t^o = F_t^{Ao} U_o + T_t^{Ao}$$

$$\hat{U}_t = \hat{F}_t^A U_o + \hat{T}_t^A$$

$$\hat{U}_t = \hat{F}_t^{Ac} U_t^o + (\hat{T}_t^A - F_t^{Ac} T_t^{Ao})$$

[Unintentional motion]

$$\theta_t^c = \theta_t^o - \hat{\theta}_t$$

$$T_t^c = \hat{T}_t^A - F_t^{Ac} T_t^{Ao} = \hat{T}_t - \begin{bmatrix} \cos\theta_t^c & -\sin\theta_t^c \\ \sin\theta_t^c & \cos\theta_t^c \end{bmatrix} T_t^o$$

[Compensated image frame]

$$I_t^c \begin{bmatrix} x \\ y \end{bmatrix} = I_t\left(\begin{bmatrix} \cos\theta_t^c & -\sin\theta_t^c \\ \sin\theta_t^c & \cos\theta_t^c \end{bmatrix}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} T_t^{cx} \\ T_t^{cy} \end{bmatrix}\right)\right)$$

04 Experimental Results and Analysis

Sample image

(a) Island (b) Road (c) University

(a) : small motions less than 2 pixels and high frequent jiggles.
(b) : small motion less than or equal to 1 pixel and irregular conditions.
(c) : small motions less than 2 pixels and high frequent rotational and translational jiggles.

Parameter

$r = 0.1,\ N_0 = 100, N_{thresh} = 5,\ \Delta T_{thresh} = 10,\ \Delta T_p = 10,\ p = 20$

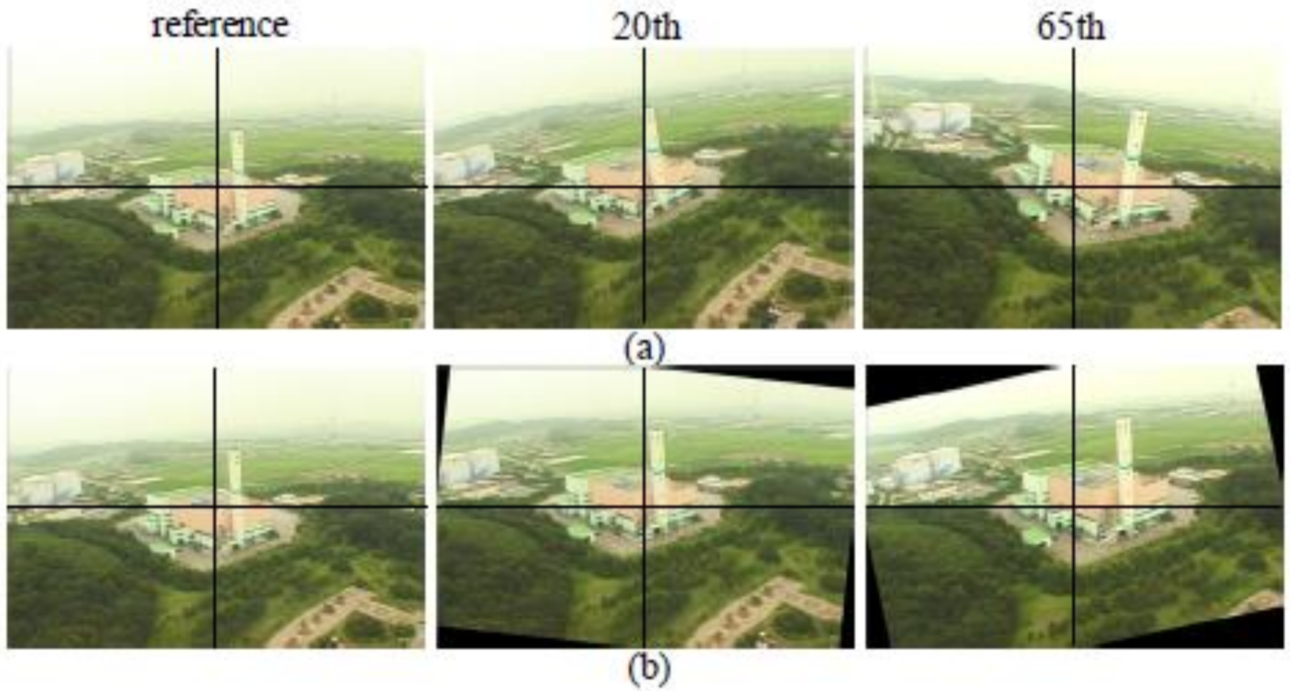
Adaptive KF

$\Delta T_p = 10,\ CT_1 = 1,\ CT_2 = 3,\ CT_3 = 5,\ CT_4 = 7,\ R_{t1} = 4R_{t0},\ R_{t2} = 20R_{t0},\ R_{t3} = 40R_{t0},\ R_{t4} = 80R_{t0}$

where R_{t0} depends on specific applications and is usually set as $R_{t0} = 1.0$

04 Experimental Results and Analysis

☆ Example



(a) : Original image frames
(b) : Stabilized image frames

☆ Effectiveness of Motion Prediction With KF

COMPARISON OF TRACKING SPEED IN FPS UNIT

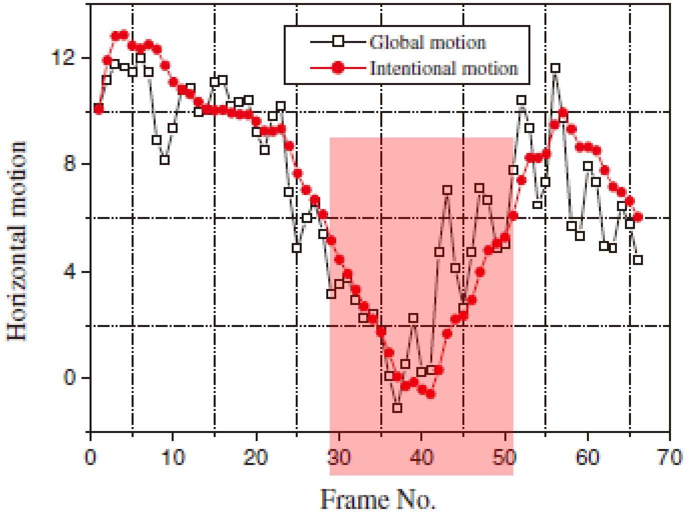
Video Seq.	With MP	Without MP
Island	33.33	31.66
Road	31.66	31.66
University	31.94	30.55

10 times
avg : over 100 frames

04 Experimental Results and Analysis

☆ Performance of Adaptive KF

(a)

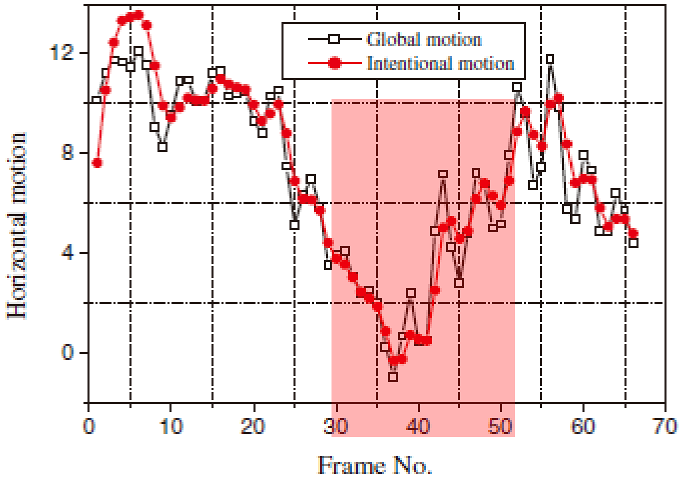


(a) Island

(1) high-frequency jiggles exist in the horizontal direction

(2) the amount of jiggles between the 30 and 50th frames changes frequently

(b)



(a) : Performance with adaptive KF
(b) : Performance without adaptive KF

04 Experimental Results and Analysis

Computational Complexity

COMPUTATION TIME IN MS/FRAME UNIT

Video Seq.	100 frames	150 frames	210 frames
Island	30.0	33.3	38.1
Road	31.6	33.3	38.0
University	31.3	40.0	39.9

COMPUTATION TIME FOR DIFFERENT PARTS IN MS UNIT

Video Seq.	FP-Sel	C-GIM	C-UM
Island	108.0	8.6	28.8
Road	112.1	8.5	29.0
University	109.9	10.5	28.8

PROCESSING SPEED FOR VIDEO SEQUENCES WITH DIFFERENT RESOLUTION

Resolution	Pixel ratio	C-GIM (ms)	C-UM (ms)	Speed (fps)
720×480	4.5	8.6	28.8	26.23
640×480	4	10.0	24.3	29.08
320×240	1	2.8	6.5	106.78

05 Conclusion

☆ We proposed a **new DIS algorithm** which can obtain a good stabilized performance in **real time**.

(KLT tracker + Kalman filter)

☆ The intentional motion is obtained with the proposed **adaptive Kalman filter**

☆ The developed algorithm has **better robustness against irregular conditions** than the conventional ones.

Q & A