ONLINE TRAINING-ORIENTED VIDEO SHOOTING NAVIGATION SYSTEM BASED ON REAL-TIME CAMERAWORK EVALUATION

M.Kumano

Ryukoku University
1-5 Yokotani, Seta Oe-cho, Otsu
520-2194, Japan
kumano@rins.ryukoku.ac.jp

K.Uehara and Y.Ariki

Kobe University
1-1 Rokkodai, NADA, Kobe
657-8501, Japan
uehara@kobe-u.ac.jp

ABSTRACT

In this paper, we propose an online training-oriented video shooting navigation system focused on camerawork based on video grammar by real-time camerawork evaluation to train users shooting nice shots for the later editing work. In this system, the processing speed must be very high so that we use a luminance projection correlation and a structure tensor method to extract the camerawork parameters in real-time. From the results of camerawork analysis, the results of each frame are classified into 7 camerawork types and the system issues 6 types of alarms and navigates users along the specified shot depending on camerawork based on video grammar in real-time while shooting the shot. Thereby, users can naturally acquire shooting style by trying to decrease alarms of improper camerawork without a consideration of the video grammar.

1. INTRODUCTION

In last years, low-priced digital video cameras spread widely and general users can shoot a video freely. Moreover, there are many cases where amateurs' videos are used for broadcasting. However, the video shooting by a novice has several problems concerning improper camerawork such as "Hand shake", "Too fast motion", "Speed fluctuation", "Rapid speed", "Too sharp curve" and "Serpentine motion" in contradiction to the special rules called "video grammar"[1]. The video grammar is composed of rules to extract appropriate shots and to connect them such as "A panning or zooming shot must follow and be followed by 1 second fixed shot", "The tempo of the pan and the zoom must be constant" or "The movement of the pan and zoom should be stable". In order to make these rules applicable, the metadata such as camerawork included in the shots have to be extracted with accuracy. We have developed an off-line video shooting navigation system[2]. Although the system decreased the problem of the camerawork, the number of the shots which reflects the video grammar has not been increased through the system. Moreover, in this system, the application of video grammar even had improper camerawork; "Too fast motion". To solve this problem, we propose an online training-oriented video shooting navigation system as the first step to the novice. The system instructs the form to the user, which includes two shot types depending on the camerawork consisting of only fixed shot or partial fixed and camerawork shot of 6 moving types in Fig.1. In this system, the processing speed must be very high so that we use a luminance projection correlation[3] to extract the camerawork parameters in real-time and a modified method of a structure tensor histogram[4] secondarily for accuracy improvement.

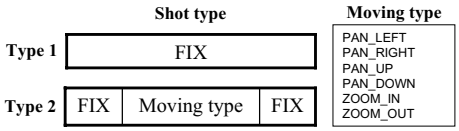

Fig. 1. Shot types as the form depending on camerawork.

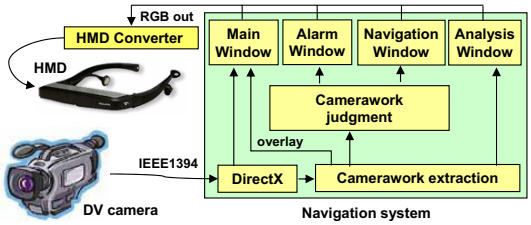

Fig. 2. Flow in the suggested system.

2. VIDEO SHOOTING NAVIGATION SYSTEM

2.1. Processing flow

Fig.2 shows the processing flow of the training-oriented video shooting navigation system. At first, the input image sequence from the DV camera is scaled down in half size by DirectX via IEEE1394 and is displayed in the "Main Window" as an alternative of the viewfinder of the DV camera. Then four amounts of camerawork at frame f such as $Pan_{lr}^{HLP}(f)$ in left-right direction using HLP(horizontal luminance projection), $Pan_{ud}^{VLP}(f)$ in up-down direction using VLP(vertical luminance projection), Zio_f^{HLP} and Zio_f^{VLP} in zoom in-out[2, 5] are independently calculated using the luminance projection correlation method[3] and these amounts are displayed in the "Analysis Window".

However, the input image sequence has time fluctuation due to asynchronous access in DirectX and a calculation cost of the camerawork extraction and judgment process. Thus, it is necessary to normalize these amount of the camarawork by the time interval between the current and the previous time in the input image sequence. Also, the strength $S_{pan}(f)$ and the direction $\theta_{pan}(f)$ calculated by both $Pan_{lr}^{HLP}(f)$ and $Pan_{ud}^{VLP}(f)$ as components of a panning vector, and the amount of $Zio(f)$ calculated by both Zio_f^{HLP} and Zio_f^{VLP} is normalized by time interval and dis-

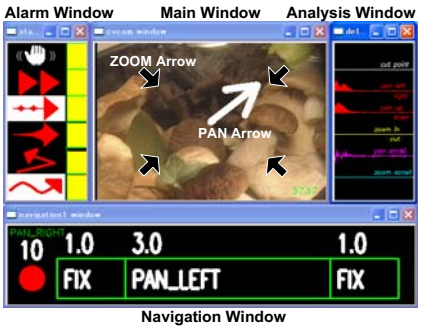

Fig. 3. Display of the suggested system.

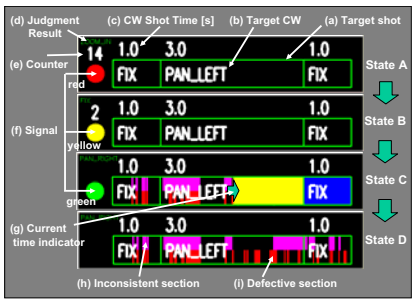

Fig. 4. Process of the navigation window.

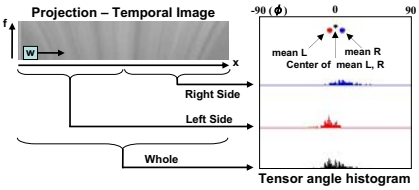

Fig. 5. Modified method of the structure tensor method.

played by overlaying as a white allow(Pan) or grouped black allows(Zoom) of vector quantity as shown in "Main Window" of Fig.3. Next, each improper camerawork is displayed in the "Alarm Window". Also, the camerawork specified for the user in advance in the "Navigation Window" is compared with the estimated camerawork parameters such a right-pan, left-pan, up-pan, down-pan, zoom-in and zoom-out in the camerawork judgment process.

Finaly, the user can take a look of the four windows via the HMD(Head mount display) simultaneously. By trying to decrease alarms of improper camerawork using this system repeatedly, the user can naturally acquire a shooting style without consideration of the video grammar.

2.2. Alarm window

Six icons in "Alarm Window" are shown in Fig.3, each of which corresponds to "Hand shake", "Too fast motion", "Speed fluctuation", "Rapid speed", "Too sharp curve" and "Serpentine motion" respectively. If any of the improper camerawork happens, the background color of the corresponding icon reverses from black to white and the neighboring box shrinks in conjunction with it.

2.3. Navigation window

Fig.4 shows the transition of 4 states(State A - State D) on the "Navigation Window". (a) is a target shot based on video grammar. In this case, the shot consists of 3 partial camerawork sections. Also, (b) is a camerawork of these partial sections such as `FIX`, `PAN_RIGHT`, `PAN_UP`, `PAN_DOWN`, `ZOOM_IN`, `ZOOM_OUT`. (c) is the duration in seconds corresponding to the partial section.

The result of the camerawork judgment on each frame is presented in (d) area in real-time. (e) is the countdown timer and the circle of (f) is a signal with red, yellow and green indicating the remaining time. The user tries to plan the shooting between State A and State B, and the navigation is started when the countdown timer becomes zero. The pointer (g) indicates the exact current time location of the shot on the State C. The user tries to trace the target shot according to the position of the current time while shooting the shot. In the State C, if the result of camerawork judgement at (d) is different from the specified camerawork at the current position, the inconsistent sections are painted out with pink in the upper half of the target shot such as (h). Also, the sections of the extracted improper camerawork(defective section) are painted out with red in the lower half of the target shot such as (i). In this system, the purpose imposed on the user is to decrease the inconsistent sections and defective sections.

3. CAMERAWORK JUDGMENT

3.1. Camerawork extraction

The $S_{pan}(f), \theta_{pan}(f)$ and $Zio(f)$ based on the luminance projection correlation method[3] can extract sensitively the movement of camerawork. In this method, however, few effects can be caused by excessive detection of zooming parameter, especially when the panning happens. In this paper, we apply secondarily the modified structure tensor method[4] to the zooming extraction for accuracy improvement.

The local orientation ϕ based on structure tensor[4] of a window w in the luminance projection-temporal image shown in the upper left in Fig.5 is divided into ϕ_L in the left half side and ϕ_R in the right half side of the projection-temporal image as shown in Fig5. In the panning, $\hat{\mu}_L(\phi_L)$ in Eq.(1) approximately corresponds to $\hat{\mu}_R(\phi_R)$ in Eq.(2). However, there are different in the zooming (the measure is Eq(3)). Also, $\hat{\mu}_L(\phi_L)$ and $\hat{\mu}_R(\phi_R)$ become nearly zero (the measure is Eq(4)) or Zio_f^{HLP} and Zio_f^{VLP} are consistent in the zooming. Therefore, $Zio(f)$ is redefined as Eq.(5) by constraint conditions for accuracy improvement and the interval of the current time t_f and previous time t_{f-n_z} for time fluctuation.

$$\hat{\mu}_L(\phi_L) = mean(\phi_L) \quad (\mu-3\sigma_L<\phi_L<\mu+3\sigma_L) \tag{1}$$

$$\hat{\mu}_R(\phi_R) = mean(\phi_R) \quad (\mu-3\sigma_R<\phi_R<\mu+3\sigma_R) \tag{2}$$

$$\mu_{diff} = |\hat{\mu}_L(\phi_L) - \hat{\mu}_R(\phi_R)| \tag{3}$$

$$\mu_{center} = |\hat{\mu}_L(\phi_L) + \hat{\mu}_R(\phi_R)|/2 \tag{4}$$

$$Zio(f) = \begin{cases} 0 & (\mu_{diff}^{HLP}<\theta_{zd} \vee \mu_{diff}^{VLP}<\theta_{zd}) \\ 0 & (\mu_{center}^{HLP}>\theta_{zs} \vee \mu_{center}^{VLP}>\theta_{zs}) \\ 0 & (Zio_f^{HLP}=0 \vee Zio_f^{VLP}=0) \\ 0 & (Zio_f^{HLP}>0 \wedge Zio_f^{VLP}<0) \\ 0 & (Zio_f^{HLP}<0 \wedge Zio_f^{VLP}>0) \\ \frac{Zio_f^{HLP}+Zio_f^{VLP}}{2(t_f-t_{f-n_z})} & otherwise \end{cases} \tag{5}$$

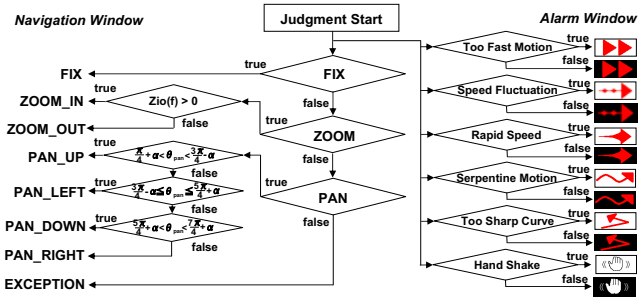

Fig. 6. Process of camerawork judgment.

3.2. Camerawork label jedgment

Fig.6 shows the process of camerawork label judgment. First, when one of the conditions $S_{pan}(f) = 0$ or $Zio(f) = 0$ is satisfied, the motion is judged as `FIX`. Next, when a condition $Zio(f) > 0$ satisfied, it is `ZOOM_IN`. Inversely, $Zio(f) < 0$ indicates `ZOOM_OUT`. Finaly, when the condition $S_{pan}(f) \neq 0$ is satisfied, θ_{pan} is classified into four direction; `PAN_UP`, `PAN_LEFT`, `PAN_DOWN`, `PAN_RIGHT` by the boundary of angles such as a $\pi/4$, $3\pi/4$, $5\pi/4$, $7\pi/4$. These labels are displayed on the "Navigation Window" as shown in Fig.4(d). In contrast, six types of improper camerawork which are displayed on the "Alarm Window" are independently calculated respectively.

3.3. Hand shake judgment

The "Hand Shake" HS_f at frame f is defined as small amounts of panning wiggled in isolation. Therefore, when the previous amount of panning is zero and the current amount is smaller than a threshold on $hs_f^{lr} = 1$ (Eq.(6)) or $hs_f^{ud} = 1$ (Eq.(7)), the motion is judged by Eq.(9) based on Eq.(8) using hand shake existence rate $\bar{hs}_f^{lr} = \sum_{n=0}^{N-1} \frac{hs_{f-n}^{lr}}{N}$ and $\bar{hs}_f^{ud} = \sum_{n=0}^{N-1} \frac{hs_{f-n}^{ud}}{N}$ within a window N.

$$hs_f^{lr} = \begin{cases} 1 & (Pan_{lr}(f-1)=0 \wedge 0<Pan_{lr}(f)<\theta_{hs}^{lr}) \\ 0 & otherwise \end{cases} \quad (6)$$

$$hs_f^{ud} = \begin{cases} 1 & (Pan_{ud}(f-1)=0 \wedge 0<Pan_{ud}(f)<\theta_{hs}^{ud}) \\ 0 & otherwise \end{cases} \quad (7)$$

$$hs_f = \begin{cases} \bar{hs}_f^{lr} & (\bar{hs}_f^{lr} \geq \bar{hs}_f^{ud}) \\ \bar{hs}_f^{ud} & (\bar{hs}_f^{lr} < \bar{hs}_f^{ud}) \end{cases} \quad (8)$$

$$HS_f = \begin{cases} true & (\theta_{hs}^{min} < hs_f < \theta_{hs}^{max}) \\ false & otherwise \end{cases} \quad (9)$$

3.4. Too fast motion judgment

When the $S_{pan}(f)$ or the absolute value of $Zio(f)$ is within thresholds in Eq(10), the motion is judged as the "Too Fast Motion".

$$TFM_f = \begin{cases} true & (\theta_{tfm}^{min} < S_{pan}(f) < \theta_{tfm}^{max}) \\ true & (\theta_{tfm}^{min} < |Zio(f)| < \theta_{tfm}^{max}) \\ false & otherwise \end{cases} \quad (10)$$

3.5. Speed fluctuation judgment

When the camerawork is smooth, the difference of the camerawork amount in consequtive frames is small. Therefore, each sf_f^p(Eq.(13)) or sf_f^z(Eq.14) using $s\bar{f}_f^p$(Eq.(11)), $s\bar{f}_f^z$(Eq.(12)) respectively becomes small. In fact, the "Speed Fluctuation" is judged by SF_f when sf_f^p or sf_f^z is larger than thresholds in Eq.(15).

$$s\bar{f}_f^p = \frac{1}{N}\sum_{i=0}^{N-1} |S_{pan}(f-i) - S_{pan}(f-i-1)| \quad (11)$$

$$s\bar{f}_f^z = \frac{1}{N}\sum_{i=0}^{N-1} |Zio_{f-i} - Zio_{f-i-1}| \quad (12)$$

$$sf_f^p = \sum_{i=0}^{N-1} ||S_{pan}(f-i) - S_{pan}(f-i-1)| - s\bar{f}_f^p| \quad (13)$$

$$sf_f^z = \sum_{i=0}^{N-1} ||Zio_{f-i} - Zio_{f-i-1}| - s\bar{f}_f^z| \quad (14)$$

$$SF_f = \begin{cases} true & (\theta_{sf} < sf_f^p \vee \theta_{sf} < sf_f^z) \\ false & otherwise \end{cases} \quad (15)$$

3.6. Rapid speed judgment

The "Rapid speed" is judged by RS_f when the acceleration of the $S_{pan}(f)$ or the $Zio(f)$ are larger than the thresholds in Eq.(16).

$$RS_f = \begin{cases} true & (\theta_{rs} < \frac{S_{pan}(f)-S_{pan}(f-1)}{t_f - t_{f-1}}) \\ true & (\theta_{rs} < \frac{Zio(f)-Zio(f-1)}{t_f - t_{f-1}}) \\ false & otherwise \end{cases} \quad (16)$$

3.7. Too sharp curve judgment

We define $\Delta P_f^\theta = \theta_{pan}(f) - \theta_{pan}(f-1)$. "Too Sharp Curve" is judged by TSC_f in Eq.(18) using $\Delta\hat{P}_f^\theta$ converted into the angle distance in Eq.(17).

$$\Delta\hat{P}_f^\theta = \begin{cases} \Delta P_f^\theta - 2\pi & (\pi < \Delta P_f^\theta) \\ \Delta P_f^\theta & (-\pi < \Delta P_f^\theta \leq \pi) \\ \Delta P_f^\theta + 2\pi & (\Delta P_f^\theta \leq -\pi) \end{cases} \quad (17)$$

$$TSC_f = \begin{cases} true & (|\Delta\hat{P}_f^\theta| > \theta_{tsc}^t \wedge S_{pan}(f) > \theta_{tsc}^r) \\ false & otherwise \end{cases} \quad (18)$$

3.8. Serpentine motion judgment

We define $k_f^p = 1$ $(\Delta\hat{P}_f^\theta > \theta_{sm}^k)$ and $k_f^n = 1$ $(\Delta\hat{P}_f^\theta < -\theta_{sm}^k)$. When the panning shake is greater than the threshold in a certain direction of right and left, $sm_f^p = \sum_{n=0}^{F-1} \frac{k_{f-n}^p}{F}$ and $sm_f^n = \sum_{n=0}^{F-1} \frac{k_{f-n}^n}{F}$ are greater simultaneously than some threshold within a window F. Therefore, when the condition in Eq.(19) using sm_f^p and the sm_f^n is satisfied, the "Serpentine motion" is judged by SM_f.

$$SM_f = \begin{cases} true & (\theta_{sm}^p < sm_f^p \wedge \theta_{sm}^n < sm_f^n) \\ false & otherwise \end{cases} \quad (19)$$

Table 1. Results of camerawork index judgment.

Index	C	M	E	Recall	Precision
FIX	511	0	94	100.0	84.5
PAN_LEFT	447	1	1	99.8	99.8
PAN_RIGHT	317	23	22	94.8	95.0
PAN_UP	402	45	0	89.9	100.0
PAN_DOWN	438	14	4	96.9	99.1
ZOOM_IN	509	19	0	96.4	100.0
ZOOM_OUT	505	19	0	96.4	100.0
TOTAL	3229	121	121	96.4	96.4

Table 2. Results of alarm judgement.

Improper camerawork	C	M	E	Recall	Precision
Hand Shake	10	0	22	100	31
Too Fast Motion	9	1	9	90	50
Speed Fluctuation	7	3	36	70	16
Rapid Speed	10	0	23	100	30
Too Sharp Curve	6	4	30	60	17
Serpentine Motion	10	0	10	100	50

4. EXPERIMENTAL RESULTS

We implemented the training-oriented video shooting navigation system on a Pentium M(1.4GHz) personal computer. The image size becomes 360 $\times$ 240 pixels by DirectX in real-time.

4.1. Results of camerawork judgment

We have carried out the experiment of camerawork label judgment by shooting 3 times a set of seven types of shots within six seconds respectively with stable camerawork. The results are shown in Table1. In the table, C, M and E indicate the number of frames with correctly judged cameraworks, the number of frames with misjudged cameraworks and the number of frames with excessive extraction of cameraworks respectively.

where Recall=100·C/(C+M) and Precision=100·C/(C+E) are defined. The recall indicates how correctly the camerawork label are judged into the true label. On the other hand, the precision indicates how accurately the judged camerawork labels include the true labels. Both are well utilized as judgment measures and the higher value indicates the better judgment result. From the table, it can be seen that the results show an impressively high accuracy.

Table2 shows the results of improper camerawork judgment. From the table, it can be seen that the results are good except for those of the precision. The degradation is attributed to reciprocal dependence of the improper camerawork. However, it indicates certain improper camerawork misjudged into another improper camerawork. Therefore, it doesn't become a fatal problem, because the purpose of the user is essentially not different from the reduction of improper camerawork.

4.2. Subjective evaluation

We carried out subjective evaluation of how the alarm and navigation are effective for user training in video shooting by four subjects who had no knowledge about video grammar. They were required to shoot 3 times a set of three types of type1 shots and six types of type2 shots with different time duration respectively.

Table 3. Camerawork match rate and alarm rate.

Try	1	2	3
Average of camerawork match rate	72.3	70.4	77.1
Average of alarm reduction rate	71.5	70.0	75.9
Total average	71.9	70.2	76.5

Table 4. Match rate in anterior or posterior FIX section of type2.

Try	1	2	3
Average of match rate in anterior FIX section	79.4	75.8	81.3
Average of match rate in posterior FIX section	47.5	48.9	61.8
Total average	63.4	62.3	71.5

Table3 shows the result of the camerawork matched frame rate and the alarmed frame rate to the whole frame of the designated shot for each 3 times trials. Also, Table4 shows the result of the camerawork match rate of partial anterior or posterior section of the type2 shot. In shooting a shot, it is difficult to stop the camera after the camera movement section which is the posterior FIX section. The table shows that user's skill improves each repeated trials.

Comparing the two tables, it can be seen that four subjects learned how to shoot the video based on video grammar by using the online training-oriented video shooting navigation system.

5. CONCLUSION

In this paper, we proposed an online training-oriented video shooting navigation system which can evaluate the camerawork in real time and issue the alarm and navigate the users to shoot a specified shot based on the video grammar. Since the match rates and the alarm reduction rate are improved by trial repeatedly, the camerawork judgment is thought to be effective for users. Therefore, the system can improve users' skill in shooting the video. We are now planning to install the software system into a real video camera.

6. REFERENCES

[1] M.Kumano, Y.Ariki, K.Shunto, K.Tsukada: "Video Editing Support System Based on Video Content Analysis", Proc. of Asian Conference on Computer Vision (ACCV) pp.628-633 (2002).

[2] K.Uehara, M.Amano, Y.Ariki, M.Kumano: "Video Shooting Navigation System by Real-Time Useful Shot Discrimination Based on Video Grammar", Proc. of ICME2004 (International Conference on Multimedia and Expo), CD-ROM, 2004.

[3] Akio Nagasaka, Takafumi Miyatake: "Real-Time Video Mosaics Using Luminance-Projection Correlation", IEICE,Vol.J82-DII,No10,pp.1572-1580,1999. (In Japanese)

[4] Chong-Wah Ngo, Ting-Chuen Pong, Hong-Jiang Zhang, Roloand T. Chin: "Motion Characterization by Temporal Slices Analysis", In Proc. of CVPR'02, pp.768–773, (2002).

[5] M.Kumano and Y.Ariki: "Automatic Useful Shot Extraction for a Video Editing Support System", MVA2002, pp.310-313, 2002.