영상신호를 이용한 상관방식 추적기에 대한 간단한 수학적인 해석

*조 재 수, **박 동 조
*한국기술교육대학교 인터넷미디어공학부
**한국과학기술원 전자전산학부
E-mail : jaesoo27@kut.ac.kr, djpark@ee.kaist.ac.kr

A Simple Mathematical Analysis of Correlation Target Tracker in Image Sequences

*Jae-Soo Cho and **Dong-Jo Park

Abstract

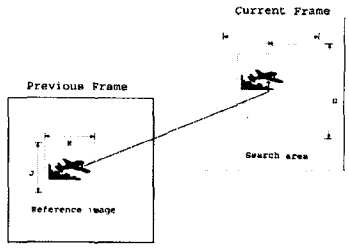
A conventional correlation target tracker is analysed with a simple mathematical approach. And, we will propose a correlation measure with selective attentional property in order to overcome the false-peak problem of the conventional methods. Various experimental results show that the proposed correlation measure is able to reduce considerably the probability of false-peaks degraded by the correlation between background images of a reference block and a distorted and noisy sensor input image.

I. Introduction

Block matching algorithms(BMA) have been widely used in the motion estimation of target tracking in addition to video coding[1]-[4]. In these approaches, the motion of a block of pixels, termed a reference block, is estimated by looking for the most similar block of pixels in the subsequent frames as shown in Fig. 1 (a).

When the reference target block does not exactly match a moving target in the search scene, the point where the reference target block best matches the target will not correspond to the center of the target. This produces an error in the correlation. If a moving target in the search scene varies greatly in size, shape, or orientation from the reference image, no true correlation peak may occur and the target in the search scene may fade into the background noise. Also, the areas in the search scene that have stronger intensity than those of the target to be tracked may produce larger correlation peaks than the true target of interest.

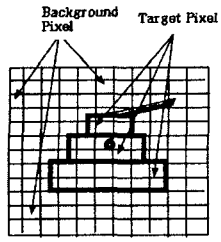
The BMA for the target tracking is somewhat different from that of motion estimation which has an important role in video coding. The usual correlation measure functions used for video coding in addition to target tracking have equal consideration to each component of the block image in the matching process. But, since the reference block of target tracking is composed of pixels of background as well as target, as shown in Fig. 1 (b), the correlation measure function in the target tracking should have different consideration according to target and background pixels in the matching process, which conform with the selective-attention property of the human visual system.



(a)

(b)

Fig. 1 Schematic diagram of the correlation tracker, (a) Target tracking by BMA, (b) Reference target block.



II. Mathematical Analysis

The reference block image $I_r(i,j)$, composed of pixels of target and background as shown in Fig. 1 (b), can be modeled mathematically as follows:

$$I_r(i,j) = t_r(i,j) + b_r(i,j) + n(i,j) \quad (1)$$
$$\simeq t_r(i,j) + b_r(i,j) \quad (2)$$

where $t_r(i,j)$ and $b_r(i,j)$ represent target and background pixels in the reference block. $n(i,j)$ is an additive noise and can be neglected for simplicity.

If a next sensor image $I_s(i,j)$, is assumed to have the previous target, the sensor image is also represented by

$$I_s(i,j) = t_s(i,j) + b_s(i,j) + n(i,j) \quad (3)$$
$$\simeq t_s(i,j) + b_s(i,j) \quad (4)$$

where $t_s(i,j)$ is a translational and geometrically distorted version of the reference target image, and $b_s(i,j)$ means an input background image which is somewhat different from the reference background image $b_r(i,j)$, due to target translation.

A cross-correlation function CCF between the reference block image and the input sensor image is computed by considering candidate blocks centered about (p,q) in the search region, and searching for the location of the best-matching block of the same size as given by

$$CCF(p,q) = \frac{1}{S_r}\sum_i\sum_j I_r(i,j)\cdot I_s(p+i,q+j) \quad (5)$$

where, $-M_1 \le p \le M_1$, $-M_2 \le q \le M_2$, $(\hat{p},\hat{q}) = arg\ \max_{(p,q)} CCF(p,q)$. S_r is the size of the reference block image, and M_1 and M_2 are predetermined integers for a search window. $(\hat{p},\hat{q})$ means an estimated target position of the current frame by the BMA. For the convenience, we will use a simple notation in the following descriptions, such as $\sum_i\sum_j = \sum_{i,j}$.

Definition 1: A target cross-correlation function $CCF_t(p,q)$ between the reference target image and the input sensor image, and a background cross-correlation function $CCF_b(p,q)$ between the reference background image and the input sensor image are defined as following:

$$CCF_t(p,q) = \frac{1}{S_r}\sum_{i,j} t_r(i,j)\cdot I_s(p+i,q+j) \quad (7)$$

$$CCF_b(p,q) = \frac{1}{S_r}\sum_{i,j} b_r(i,j)\cdot I_s(p+i,q+j) \quad (8)$$

Definition 2: A (p^*,q^*) is defined as an exact target point satisfying the following condition:

$$(p^*,q^*) = arg\ max_{(p,q)} CCF_t(p,q) \quad (9)$$

Then, following Lemma and Remark can be derived from the above definitions.

Lemma 1: The cross-correlation function, $CCF(p,q)$ between the reference block image $I_r(i,j)$ and the input sensor image $I_s(i,j)$ is represented as the sum of the target cross-correlation function and the background cross-correlation function,

$$CCF(p,q) = CCF_t(p,q) + CCF_b(b,q) \quad (10)$$

Proof:

$$CCF(p,q) = \frac{1}{S_r}\sum_{i,j} I_r(i,j)\cdot I_s(p+i,q+j)$$
$$= \frac{1}{S_r}\sum_{i,j}[t_r(i,j)+b_r(i,j)]\cdot I_s(p+i,q+j)$$
$$= \frac{1}{S_r}[\sum_{i,j} t_r(i,j)\cdot I_s(p+i,q+j)$$
$$+ \sum_{i,j} b_r(i,j)\cdot I_s(p+i,q+j)]$$
$$= CCF_t(p,q) + CCF_b(p,q).$$

Remark: The estimated target point $(\hat{p},\hat{q})$, by searching the maximum peak point of the cross-correlation function, may not be the exact target point (p^*,q^*) as following:

If $\arg\max_{(p,q)} CCF(p,q) = \arg\max_{(p,q)} CCF_t(p,q)$,

then, $(\hat{p},\hat{q}) = (p^*,q^*)$,

Otherwise, $(\hat{p},\hat{q}) \neq (p^*,q^*)$.

Lemma 1 and Remark imply well the problem of the false peaks degraded by the background

cross-correlation $CCF_b(p, q)$. Considering the correlator walk-off(or drift phenomenon), and taking count of the size changes of a moving target in the actual tracking environments, the reference block inevitably includes the background pixels near to target image.

III. Selective-attention Correlation Measure

For the precision target tracking using the block matching algorithm, a novel correlation measure is required to overcome the problems of the previous correlation trackers.

Definition 3: A weight function $w(i,j)$ is defined as a selective-attention weight satisfying the following condition:

$$w(i,j) = \begin{cases} 1, & \text{if } pixel\ (i,j) \in Target\ pixel \\ 0, & otherwise, \end{cases} \quad (11)$$

Definition 4: A selective-attention similarity measure $f_s(p, q)$, between the reference block image and the input sensor image, is defined as by considering candidate blocks centered about (p, q) in the search region,

$$f_s(p, q) = \frac{1}{S_r}\sum_{i,j} w(i,j) \cdot f(I_r(i,j), I_s(p+i, q+j)), \quad (12)$$

where, $-M_1 \le p \le M_1$, $-M_2 \le q \le M_2$, $f(\cdot)$ represents a similarity function between a pixel of the reference block $(I_r(i,j))$ and a search image $(I_s(i,j))$. M_1 and M_2 are predetermined integers for a search window.

The followings are some example of the selective-attention similarity measure for two particular similarity functions.

Example 1: A Selective-attention Cross-Correlation Function $(SCCF(p,q))$, if $f(I_r(i,j), I_s(p+i, q+j)) = I_r(i,j) \cdot I_s(p+i, q+j)$,

$$SCCF(p, q) = \frac{1}{S_r}\sum_{i,j} w(i,j) \cdot I_r(i,j) \cdot I_s(p+i, q+j)$$

$$(\hat{p}, \hat{q}) = \arg\max_{(p,q)} SCCF(p, q).$$

The correlation surface is scanned to find a maximum peak point for estimating of the target position $(\hat{p}, \hat{q})$.

Example 2: A Selective-attention Mean Absolute Difference $(SMAD(p,q))$, if $f(I_r(i,j), I_s(p+i, q+j)) = |I_r(i,j) - I_s(p+i, q+j)|$,

$$SMAD(p, q) = \frac{1}{S_r}\sum_{i,j} w(i,j) \cdot |I_r(i,j) - I_s(p+i, q+j)|$$

$$(\hat{p}, \hat{q}) = \arg\min_{(p,q)} SMAD(p, q).$$

The correlation surface is scanned to find a minimum peak point for estimating of the target position $(\hat{p}, \hat{q})$.

Lemma 2: The Selective-attention Cross-Correlation Function, $SCCF(p, q)$ is equal to the target cross-correlation function, $CCF_t(p, q)$, as following:

$$SCCF\ (p, q) = CCF_t(p, q). \quad (13)$$

Proof:

$$SCCF(p, q) = \frac{1}{S_r}\sum_{(i,j)} w(i,j) \cdot I_r(i,j) \cdot I_s(p+i, q+j)$$

$$= \frac{1}{S_r}\left[\sum_{(i,j)} w(i,j) \cdot t_r(i,j) \cdot I_s(p+i, q+j) + \sum_{(i,j)} w(i,j) \cdot b_r(i,j) \cdot I_s(p+i, q+j)\right]$$

$$= \frac{1}{S_r}\sum_{(i,j)} t_r(i,j) \cdot I_s(p+i, q+j) = CCF_t(p, q).$$

Theorem: The estimated target point $(\hat{p}, \hat{q})$ by searching the maximum peak point of the selective-attention cross-correlation function, $SCCF(p, q)$, is the exact target point, (p^*, q^*), as following:

$$(\hat{p}, \hat{q}) = (p^*, q^*) \quad (14)$$

Proof: Due to Lemma 2,

$$(\hat{p}, \hat{q}) = \arg\max_{(p,q)} SCCF\ (p, q)$$
$$= \arg max_{(p,q)} CCF_t(p, q)$$
$$= (p^*, q^*)$$

Theorem implies that if we can know exactly whether a pixel of the reference block is a target or background one, the false peaks caused by the reference background images can be removed. But, in fact, we can not know whether it is a target or background pixel. Accordingly, the exact determination of the selective-attention weight $w(i,j)$ lies at the heart of this proposed method.

VI. Experimental Results and Conclusion

Let us define a tracking contrast (TC) between target and background intensity as a criterion of difficulty for tracking environments, and an average centroid error distance $(\bar{E})$ for a precision criterion.

Definition 5: A tracking contrast (TC) measure of target image with respect to background image is defined as

$$TC = \frac{(\mu_t - \mu_b)^2}{\sigma_t^2 + \sigma_b^2} \quad (17)$$

where μ and σ are mean and standard deviations of the images, and the subscript t and b represent the target and background, respectively.

Definition 6: An average centroid error distance ($\bar{E}$) is defined as

$$\bar{E} = \frac{1}{N}\sum_{k=1}^{N}\sqrt{(i_k - \hat{i}_k)^2 + (j_k - \hat{j}_k)^2} \quad (18)$$

where N means the total frame number of the test sequence. (i_k, j_k) means the coordinate of a true target centroid and $(\hat{i}_k, \hat{j}_k)$ the one of the estimated target centroid by a tracker at the k^{th} frame.

Various kinds of image sequences were generated artificially for quantitative evaluation of the algorithm. The size of the image is 256×256 and each image includes a 30×30 rectangular target with gaussian background images. A rectangular target is moving to the right direction from an initial position drawing a sine wave trajectory.

Selection of the block size usually involves a tradeoff between two conflicting requirements. The block size should be large enough to compensate for the changes of the target size, and not to occur the correlator walk-off caused by accumulation of small tracking errors. On the other hand, it should be as small as the reference block can not include the background images near to the target.

Table 1 shows the results tracked by the *CCF* measure in various reference block size on the test sequences. By these experiments, it is obviously clear that the more the reference block image includes background pixels, the more probability of the false peaks is introduced due to the correlation between background pixels of the reference block and the input search image.

[Table 1: The average centroid error $\bar{E}$ by the *CCF* in various block sizes.]

Image /Block size	30×30	50×50	70×70
TC=2.0	0.00	0.00	1.43
TC=0.5	0.00	6.62	6.22
$TC = 0.125$	0.00	8.96	17.30
$TC = 0.025$	0.41	9.44	18.46

Tracking results by each matching criterion are also summarized for various test sequences in Table 2, where w_r means a selective-attentional wight.

[Table 2: Tracking results by various matching criteria. ($\bar{E}$ in pixels.)

Method/TC	$TC = 2.0$	$TC = 0.5$	$TC = 0.025$
CCF	0.34	9.94	16.59
SCCF with w_r	0.00	0.00	0.59
MAD	3.02	10.12	16.84
SMAD with w_r	0.00	0.12	1.17

Experimental results show that *CCF* and *MAD* are more likely to have false peaks or fail to track compared with the proposed measures. Furthermore, the tracking performance of the proposed method is somewhat independent of the reference block size. For the further study, we need to research the exact estimation of the weight for real-image sequences to adopt this selective-attention measure.

References

[1] M. C. Dudzik, "Electro-optical systems design, analysis, and testing", *The Infrared and Electro-Optical Systems Handbook*, SPIE.

[2] A. D. Hughes and A. J. Moy, "Advances in automatic electro-optical tracking systems,"' in *SPIE Vol. 1697 Acquisition, Tracking, and Pointing VI*, pp. 353-365.

[3] M. J. Chen, L. G. Chen, T. D. Chiueh, and Y. P. Lee, "A new block-matching criterion for motion estimation and its implementation," *IEEE Trans. Circ. Syst. Video Technol.*, Vol. 5, No. 3, pp. 231-236, June.

[4]. J. S. Cho, Study on Intelligent Automatic Tracking of Moving Targets in image sequences, Ph.D Thesis in KAIST, 2001.