Performance Analysis of Online Weighted Multiple Instance Learning for Single Face Tracking at Outdoor Environment

Suryo Adhi Wibowo[1], Eun Kyeong Kim[1], Eunseok Jang[1], and Sungshin Kim[2]

[1]Department of Electrical and Computer Engineering, Pusan National University, Busan, Korea
{suryo, kimeunkyeong, esjang}@pusan.ac.kr
[2]School of Electrical and Computer Engineering, Pusan National University, Busan, Korea
sskim@pusan.ac.kr

Abstract. It is rarely that researcher used children's face as an object for face tracking. Since they have a random behavior such as abrupt movement, fast movement, and pose changes, it could make tracking failed. In this paper, we will analyze online weighted multiple instance learning (WMIL) as a method for single face tracking, with face from children as an object for tracking in the outdoor environment. We used this method because the object will be represented as Haar-like feature and boosting method also included. In this paper, we analyzed parameters from online WMIL such as window size for searching the target, number of rectangles, and learning rate. Our simulation results show that the best combination of its parameter are 25, 0.70, and 6, respectively. We also compared between this method, tracking based on histogram, and point. The results shows that online WMIL produce error minimum result on an even keel.

Keywords: children's face, face tracking, online WMIL, outdoor environment.

1 Introduction

Tracking is one of important research area in the computer vision field. Many applications which can be developed by using tracking method such as surveillance applications, human-computer interaction (HCI) applications, robotics, to medical applications. The principal of tracking is given the initial state of a target object in the first frame, we should estimate the target position from the subsequents frames. The oldest approach that commonly used to represent the object is uses the histogram. This approach gives high accuracy if the background has a significant different color with the object but it will be failed if the background does not have a significantly different color with the object where this condition represent the condition in real environment.

To solve this problem, recently, the focus research in tracking field is not only using oldest approach in tracking method but they combine tracking with learning and detection which has been performed by Kalal and his colleagues [1], approaching tracking by using detection has been performed by Babenko and his colleagues [2]. They represent adaptive appearance model for the object and using online multiple instance learning (MIL) for detection instead of tracking. Zhang *et al*. [3] improved the accuracy of online MIL by using weighted for the positive samples. In this paper, we follow Zhang's work and analyzed the parameters from their method to get high precision for single face tracking at outdoor environment. As we know that there are a lot of research result about face tracking included single or multiple face tracking and research in this area was rapidly developing after Viola *et al*. [4] given the result about face detection using weak classifier. Unfortunately, the object for face tracking usually used face from adult person and it is rarely used face from children, since tracking to the children could be failed because their behavior. They can produce natural problems in tracking field such as abrupt movement, fast movement, large pose changes, occlusion, and etc. The contributions of this paper are analyzing the coefficient of combination parameters in online WMIL to get precision result for single face tracking using children's face under abrupt movement, fast movement, pose change, and cluttered background problems.

The paper is structured as follows: In Section 2, we present Online WMIL. In Section 3, data set and performance measurement are described. In Section 4, we present results and discussion. Conclusions of this paper are explained in Section 5.

2 Online Weighted Multiple Instance Learning

Basic idea for online WMIL is a tracking by detection approach. It uses MIL method where it is based on voting of boosting weak classifiers. Because weak classifier is a binary classifier, it needs two inputs which have two labeled outputs. Two inputs represent bag of the object and bag of the background. Updating the classifier in each iteration is needed. To improved the precision for tracking, Zhang [3] added the weight for positive samples where it is represent the object. Both of online MIL and WMIL are using Haar-like feature for representing the object. This feature is more robust than histogram feature and point feature for representing the object. Because of that reasons, our work is follows and based on their work [3]. Fig. 1 illustrates the basic flow of online WMIL tracking.

A good image object detection algorithm is accurate, fast, and does not require exact locations of objects [4], but the online WMIL tracking is a tracking method by detection approach. It is not a tracking method by prediction approach. So, it requires an exact location for the tracking and produce the result that have exact location is useful to make an applications. Initializing the object that will be track in the first frame is an important step for the tracking. Precision tracking will be failed if the user incorrect for giving the region of the object that will be tracked. In this paper, first, we select and give a mark the region of face which want to be tracked. After this, the selected region is extracted the feature by using integral image. The result from integral image is a haar like feature where computed by the sum of weighted pixels in

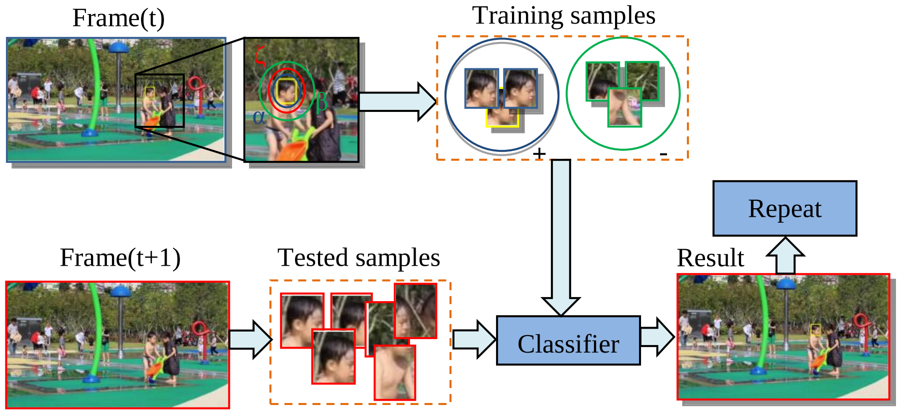

Fig. 1. Basic flow of online WMIL tracking. In the training samples box, blue and green circle represent bags of positive samples and negative samples, respectively. Image with yellow rectangles in the training samples is the most important sample. The algorithm will be repeated where frame(t) replaced to frame(t+1), frame(t+1) to frame(t+2), and etc. until last frame.

several rectangles. The number of rectangles for each sample ranges from 2 to 6. Random selection is used to select the locations of rectangles in the sample. Feature vector of each sample a is represented as $f(a)=(f_1(a),\ldots,f_K(a))^T$ and K = 15. The principal of posterior probability of labeling sample a to be positive is computed using Bayesian theorem

$$p(b=1|a)=\sigma\left(\ln\left(\frac{p(a|b=1)p(b=1)}{p(a|b=0)p(b=0)}\right)\right), \tag{1}$$

where $\sigma(z)=1/(1+e^{-z})$ is sigmoid function and $b\in\{0,1\}$ is a binary label of sample a and classifier $H_K(a)$ is the discriminative appearance model that defined as

$$H_K(a)=\ln\left(\frac{p(a|b=1)p(b=1)}{p(a|b=0)p(b=0)}\right)=\sum_{k=1}^{K}h_k(a), \tag{2}$$

where

$$h_k(a)=\ln\left(\frac{p(f_k(a)|b=1)}{p(f_k(a)|b=0)}\right). \tag{3}$$

Online WMIL works based on positive and negative samples in the positive and negative bags $\{A^+,A^-\}$, respectively. Assume we have N positive samples

$\{a_{1j}, j = 0, \ldots, N-1\}$ and L negative samples $\{a_{0j}, j = N, \ldots, N+L-1\}$ which is represent the object and background, respectively. The positive probability could be defined as follows:

$$p\left(b=1 \mid A^{+}\right)=\sum_{j=0}^{N-1} w_{j0}\, p\left(b_1=1 \mid a_{1j}\right), \tag{4}$$

where $p\left(b_1 = 1 \mid a_{1j}\right)$ is the posterior probability (i.e.,(1)) and w_{j0} is a weight

$$w_{j0} = \frac{1}{c} e^{-\left|l\left(a_{1j}\right)-l\left(a_{10}\right)\right|}, \tag{5}$$

where $l(\cdot) \in R^2$ is the location function and c is a normalization constant that output from sigmoid function computed using discriminative classifier for positive samples as an input. Eq. (5) is represent the distance between sample $a_{1,j}$ and sample a_{10}. Negative probability could be defined as follows:

$$p\left(b=0 \mid A^{-}\right)=\gamma \sum_{j=N}^{N+L-1}\left(1-p\left(b_0=1 \mid a_{0j}\right)\right), \tag{6}$$

where γ is a constant which computed from sigmoid function which is uses discriminative classifier for negative samples as an input. Because this system is based on weak classifier h_k, following criterion could be selected as an output for weak classifier:

$$h_k = \arg\max_{h \in \Phi} \left\langle h, \nabla \ell(H) \right\rangle \Big|_{H=H_{k-1}}, \tag{7}$$

where Φ is a weak classifier pool and $\left\langle h, \nabla \ell(H) \right\rangle$ is the inner product criterion for selected weak classifier. This selection criterion is more efficient than directly directly maximizes the log-likelihood function used by MIL tracker

$$h_k = \arg\max_{h \in \Phi} \ell\left(H_{K-1} + h\right), \tag{8}$$

it caused the instance probability and bag probability after selecting one weak classifier will not need to be computed, as describe in [5].

3 Data Set and Performance Measurement

In this section, we will explain about the data and performance measurement that we used. We used three data which is represent three different objects. The data represents image sequences which is recorded by using canon eos 100d. The distance between camera and the object has range from 3 m to 5 m. Each raw data has same size and color format. The size of this data is 480×854×3 which represent rows, columns, and channel, respectively. Table 1 shows the detail explanation about the data. Each frame of each data manualy labeled by us for the ground truth.

We follow Wu and his colleagues [6] for analyzing the performance which used precision as performance parameter. The precision performance calculates the distance error between the centroid of ground truth and the centroid of bounding box from detected object. Fig. 2 shows the illustration about distance error between centroid of the ground truth and the output from online WMIL. The average of precision preformances is calculated from addition precision performance every frames in one subject divided the number of frames. Three parameters that influence in online WMIL are window size for searching the target, number of rectangles, and learning rate. For the size of window searching, we analyze the size are 15 and 25. Number of rectangles that we used are 4, 5, and 6. The last parameter is learning rate. We used 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00 for learning rate parameters. In this

Table 1. Data description.

Data name	Detail problems	Number of frames
Subject 1	Abrupt movement, fast movement, cluttered background, and pose changes.	116
Subject 2	Fast movement and cluttered background.	67
Subject 3	Abrupt movement, fast movement, cluttered background, and pose changes.	110

Fig. 2. From left to right: target tracking in subject 1 (left), subject 2 (middle-left), subject 3 (middle-right), and the distance error (yellow line) between centroid of the ground truth (blue) and the output (red) for precision performance parameter (right).

paper, the positive samples N that we used is 45 and 42 is used for negative samples L.

In this paper, we also compared between the optimized online WMIL with the conventional tracking method. We used histogram-based and point-based for the conventional tracking method. For histogram-based, we used histogram as a feature for representing the object. Then, camshift algorithm will be used for tracking algorithm. For point-based, we used point as a feature for representing the object. Then, we used Kanade-Lukas-Tomasi KLT algorithm [7] as a tracking algorithm.

4 Results and Discussion

In this section, results and discussion explained. Table 2, Table 3, and Table 4 represent the average distance error from subject 1, subject 2, and subject 3, respectively. In these table, we can see clearly that the combination of three parameters from online WMIL where is learning rate, searching window size, and number of

Table 2. Average distance error (pixels) in subject 1.

Search window size	15			25		
Number of rectangles / Learning rate	4	5	6	4	5	6
0.70	19.79	21.50	**19.79**	43.59	**18.92**	**8.64**
0.75	23.12	**19.63**	22.45	**22.11**	40.97	20.38
0.80	19.05	32.19	23.95	45.92	42.49	49.60
0.85	**17.31**	27.03	22.24	46.81	46.61	44.77
0.90	35.72	44.29	43.55	45.45	49.12	48.68
0.95	25.18	33.33	29.14	45.71	41.46	40.93
1.00	375.64	375.69	376.24	386.92	386.84	387.22

Table 3. Average distance error (pixels) in subject 2.

Search window size	15			25		
Number of rectangles / Learning rate	4	5	6	4	5	6
0.70	**35.46**	**36.27**	42.12	**3.97**	3.23	**3.11**
0.75	42.01	38.97	**40.70**	4.87	**3.12**	3.46
0.80	36.77	42.39	41.27	3.86	19.67	3.50
0.85	39.57	41.81	48.81	2.96	20.81	3.45
0.90	40.04	63.87	40.73	3.74	43.84	6.35
0.95	59.76	36.36	46.17	112.67	192.76	44.98
1.00	387.89	387.17	388.25	400.85	400.45	399.11

Table 4. Average distance error (pixels) in subject 3.

Search window size	15			25		
Learning rate \ Number of rectangles	4	5	6	4	5	6
0.70	7.39	13.30	7.25	**5.56**	9.52	5.93
0.75	7.61	6.33	16.30	6.64	10.69	5.48
0.80	7.63	5.04	8.08	7.74	**8.93**	7.36
0.85	**4.13**	**3.57**	12.04	11.08	21.05	5.12
0.90	5.15	7.48	**7.13**	9.37	27.72	6.37
0.95	11.04	10.53	7.27	29.17	15.21	**4.02**
1.00	317.89	317.33	318.42	364.41	364.77	363.53

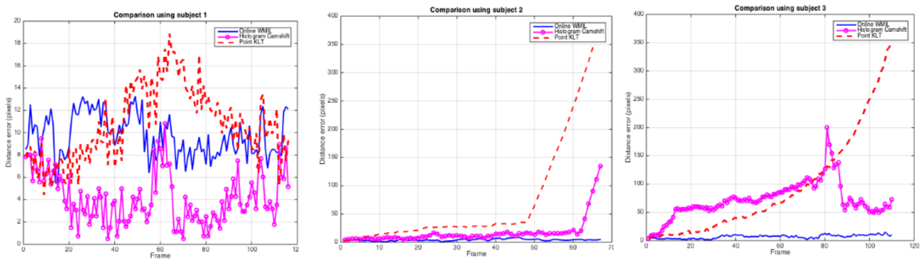

Fig. 3. Left to the right: the distance error (pixels) comparison between online WMIL (——, blue), histogram camshift (-o-, magenta), and point KLT (----, red) for Subject 1 (left), Subject 2 (middle), and Subject 3 (right), respectively. In these picture, vertical axes and horizontal axes represent distance error (pixels) and frame, respectively.

rectangles is equal to 0.70, 25, and 6, respectively, always has stable performance which is produce minimum average distance error for each subject. The problems such as blurred image from abrupt and fast movement, cluttered background, and pose changes can be handle by using these coefficient. Bold mark in the tables represent minimum average distance error for each number of rectangles group.

Fig. 3 shows the comparison between online WMIL tracker, histogram camshift, and point KLT. Online WMIL can produce minimum error on an even keel. This is because Haar-like feature more robust to handle the problems such as abrupt movement, fast movement, pose changes, and cluttered background. Furthermore, online WMIL used integral image and boosting scheme for classification, this approach can produce fast result. For histogram feature, it is good for tracking if the object and the background have the color significantly different. Unfortunately, in the real condition, sometimes the color between the object and the background looks similar and illumination also will make an effect for the color and this is the drawback for histogram feature. We can see from Fig. 3 that suddenly, histogram camshift produce big error (tracking failures) because the color between the object and the background is similar. The last tracking algorithm in this comparison is point KLT algorithm. Point KLT algorithm used a point to represent the object. Even the KLT algorithm is one of great tracking algorithm, but using point feature sometimes is not appropriate for tracking problems. It is because sometimes point feature could not handle the tracking problems such as fast movement and pose changes.

5 Conclusions

We have implemented online WMIL method for single face tracking which use face from children at outdoor environment and also analyze the parameters of this method. From three parameters that have been analyzed by us, the combination parameters producing stable high precision are: learning rate, searching window size, and number of rectangles are 0.70, 25, and 6. Problems in our case such as fast movement, cluttered background, and pose changes could be handled by these parameters. We also compared the online WMIL with the conventional tracking algorithm such as histogram camshift and point KLT. Based on our simulation result, the online WMIL can produce error minimum on an even keel.

Acknowledgements. This work was supported by BK21PLUS, Creative Human Resource Development Program for IT Convergence.

References

1. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 34(7), 1409-1422 (2012)
2. Babenko, B., Yang, M-H., Belongie, S.: Visual Tracking with Online Multiple Instance Learning. In: IEEE conference on Computer Vision and Pattern Recognition, pp. 983-990. IEEE Press, Miami (2009)
3. Zhang, K., Song, H.: Real-time Visual Tracking via Online Weighted Multiple Instance Learning. 46, 397-411 (2013)
4. Viola, P, A., Jones, M, J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: IEEE conference on Computer Vision and Pattern Recognition, pp. 511-518. IEEE Press (2001)
5. Wang, Q., Chen, F., Xu, W., Yang, M-H.,: An experimental comparison of online object tracking algorithms. In: Proceedings of SPIE: Image and Signal Processing Track. 8138. (2011)
6. Wu, Y., Lim, J., Yang, M-H.: Object Tracking Benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence. (1), pp.1, PrePrints, doi:10.1109/TPAMI.2014.2388226
7. Jianbo, S., Tomasi, C.: Good Features to Track. In: IEEE conference on Computer Vision and Pattern Recognition, pp. 593–600. IEEE Press (1994)