Online Determination of Track Loss Using Template Inverse Matching

Rong Liu, Stan Z. Li, Xiaotong Yuan, Ran He

Online Determination of Track Loss Using Template Inverse Matching

Rong Liu, Stan Z.Li, Xiaotong Yuan, and Ran He
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation,Chinese Academy of Science
100190,Beijing,P.R.China
{rliu,szli,xtyuan,rhe}@cbsr.ia.ac.cn

Abstract

Online determination of track loss in the absence of ground truth is an important and challenging problem in visual tracking systems. In this paper, we construct a novel track loss determination strategy using Template Inverse Matching (TIM). The idea of TIM is to inverse the common process of template tracking match. Ground truth is not required to be known. It is proved to be highly efficient and accurate, and is adaptive to all the visual tracking frameworks. The proposed strategy is justified in the theoretical framework of Stable Marriage (SM) problem. In this paper, we prove that the combination between the target position got from the TIM algorithm and the original one can meet the constraints of SM pairs and is irreplaceable when a correct tracking process is performed. This guarantees the stability of object tracking. Compariative experiments show high accuracy of the present method in determining track loss. The performance of tracking systems can be improved based on online track loss determination.

1. Introduction

In video applications, continuously tracking an object over a period of time to generate a persistent trajectory is important. Visual tracking has received wide attention in computer vision. Not only is it a core engine in building an intelligent video analysis system, but also it can be utilized as a basis for developing high level intelligent analysis systems. However, no tracking algorithms can perfectly perform in all conditions due to crowded environments, clutter, changing illumination and self-occlusion are ubiquitous, , and track loss is inevitable in real tracking systems. In this context, online determination of track loss is important for practical visual tracking systems: (1) With online determination of track loss, systems can do some adjustments, such as reset, by itself to avoid sustaining affection caused by just once track failure. (2) Based on track loss determination, tracking systems can update tracking strategy online, such as changing the feature template, to improve the accuracy of tracking results. For these reasons, the approach for online determining track loss can play an important role in visual tracking systems. Here, online means that the determination is automatic, without use of any ground truth, and that it is also a causal determination method.

One possibility is to directly use the similarity of track matching to determine track loss, and the weak similarity correspond to a track failure (called Weak Similarity Determination (WSD) hereafter). WSD is not a very credible approach to determine track loss. Because the target with weak similarity can be also well tracked, when the background's similarity is much weaker than the target's. For the same reason, the target with strong similarity is also able to lose track, when the background's similarity is as strong as the target's.

Another possibility, called Similarity Covariance Determination (SCD), takes advantage of covariance of similarity to determine track loss and this could improve over WSD. While the target is kept in track, the similarity distribution of track matching always has a sharp peak, otherwise the distribution is almost homogeneous. It is more reliable to take advantage of similarity covariance to determine track loss in SCD than to directly make use of similarity value in WSD. In the robot visual system of [8], the SCD method is adopted. But in the real world, the SCD is always disturbed by peaks of similarity distribution appearing in background in clutter environment. In [10], an entropy based criterion is used to evaluate the statistical characteristics of the tracked density function. It is similar to the SCD approach.

Under specific tracking frameworks, some other methods could also realize online determination of track loss. But since they are limited to specific tracking frameworks, it is hard to extend them to more generalized tracking systems. In [13] [14], the track loss determination strategy is deduced from the kalman filter structure. The work in [4] [3] is specific to contour based tracking systems. In [15], Wu *et al.* propose a performance evaluation strategy for tracking systems based on particle filter using a time reversed Markov chain.

In [16], the approach combines various tracking performance evaluation strategies, such as trajectory complexity, motion smoothness, scale constancy, shape and appearance similarity, to determine track loss. This kind of method needs to compute many estimation indexes, so it is time consuming. The utilization of motion information in this method limits it to static camera systems.

In this paper, we propose a novel track loss determination strategy named Template Inverse Matching (TIM). Finding the region which matches the previous templates as closely as possible in current image is a common way for position the target in visual tracking. The TIM, which takes advantage of current target as the template to match that in the last frame, is an inverse process of common template tracking match. To determine the track loss, the result obtained from the proposed algorithm is compared with the target position in last frame. Similar positions reflect a weak disturbance from the background and a more correct tracking result. In contrast, a strong disturbance can be in the tracking process and lead to a probably failing result. The proposed strategy is online, without use of any ground truth, and can be adaptive to all the visual tracking frameworks. Experiments demonstrate its good performance.

It is worth mentioning that the approach proposed in [15] which is based on time reversed Markov chain also make use of the thought of inverse tracking from current frame to the last one. But there are two essential distinctions: 1) The method presented in [15] is based on the particle filter framework, whereas our approach is not specific to any certain tracking framework. This means that our approach can adapt to more generalized tracking algorithms, such as Mean Shift, Particle Filter and so on; 2) The method proposed in [15] is deduced from the Markov chain, whereas our approach is justified by the theoretical framework of Stable Marriage (SM) problem [6].

The concept of SM is first introduced by Gale and Shapley in [6], and is used to find a set of stable matching pairs between two sets. Gale and Shapley show that if the sets are disjoint then at least one set of stable marriages exits. In general, there are several such sets of stable marriages. In [12], McVitie and Wilson propose an algorithm to find all of these sets. In recent years, many researches [5] [7] [9] focus on this topic. The estimated positions of tracking targets in last and current frames can be considered as two members in different sets. If the pair of them is stable in each possible sets of stable marriages, the tracking result can be considered as reliable. In this paper, it is proved that the combination between the target position got from TIM and the original one can meet the constraints of SM pairs and is irreplaceable when a correct tracking process is performed. So the framework of SM problem provides an important theoretical support to the proposed track loss determination strategy.

Taking advantage of online determination of track loss, the performance of tracking systems can be improved through online updating the tracking strategy. For this purpose, experiments using proposed track loss determination approach are provided.

The reminder of the paper is organized as follows. In section 2, the online track loss determining strategy using TIM is presented first, and then it is proved to be reasonable through the theoretical framework of Stable Marriage (SM) problem. In section 3, a method to improve the performance of tacking systems based on track loss determination is provided. Some experimental results and conclusions are presented in section 4 and 5.

2. Track Loss Determination Using TIM

In this section, a novel track loss determination strategy using Template Inverse Matching (TIM) is introduced, and its theoretical support from the SM problem is formulated. When a correct tracking process is performed, it can be proved that the combination between the target position got from the TIM algorithm and the original one can meet the constraints of SM pairs. Furthermore, they are the exclusive SM pair existing in all the possible sets of stable marriages. This gives us a criterion to distinguish between track loss and track success.

2.1. The TIM Approach

Finding the region which matches the previous templates as closely as possible in current image is a common way for position the target in template visual tracking. Assuming an inverse of the common template matching approach, we take advantage of current target as the template to match the last frame. Under intuitive consideration, it is reasonable to expect that the results of above template inverse matching may reflect some tracking performance as follows. If the target position obtained from the template inverse matching is consistent with the original one in last frame, it means that the estimated target in current frame coincides with the original target. So the result of tracking can be considered as creditable. If the result of template inverse matching locate in the region of background, it means the tracking is disturbed by the background, and the track loss is possible to take place. This is the motivation of TIM. Its theoretical support will be given in the following subsection.

Let I_n be the nth frame of the video sequences, and T_n be the target in the nth frame. The target tracking process can be described as below:

$$T_n = M(I_n, T_{n-1}) \tag{1}$$

The M is denoted as a kind of tracking algorithm, such as Mean Shift, Particle Filter and so on. After inputting current frame and target in last frame, we can estimate the target in current frame through tracking algorithms M. The TIM is an inverse of the above process:

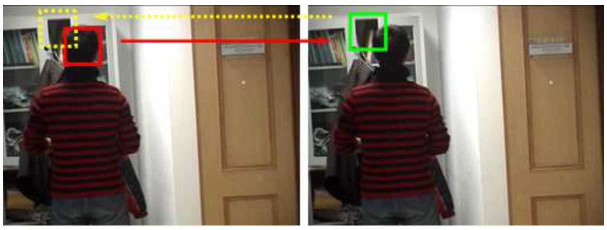
$$T'_{n-1} = M(I_{n-1}, T_n) \tag{2}$$

T'_{n-1} is the target in the $n-1$th frame estimated by TIM. To determine the track loss, the squared distance d_n between the target center of last frame and the one got from TIM is measured. Their distance is scaled by height and width of the target for constructing a uniform determination criterion. The expression of d_n is shown as below:

$$d_n = \sqrt{\left(\frac{C_x(T_{n-1}) - C_x(T'_{n-1})}{W(T_{n-1})}\right)^2 + \left(\frac{C_y(T_{n-1}) - C_y(T'_{n-1})}{H(T_{n-1})}\right)^2} \tag{3}$$

where the $[C_x, C_y]$, H and W are respectively center coordinate, height and width of the target. If d_n is greater than a given threshold, the track loss in I_n is determined. In Figure 1, two kinds of template matching results got through TIM and their corresponded track loss determining are shown.

(a)



(b)

Figure 1. The match results of TIM: (a) a matched result, and (b) an unmatched result. The image on left is last frame, and the one on right is current frame. The real red and green rectangle respectively show the tracking results in last frame and current frame. The dashed yellow rectangle shows the result of TIM. The real red arrow shows the common tracking direction from last frame to current frame, and the dashed yellow arrow denotes the direction of TIM.

In Figure 1(a), there is a short distance between the target center of last frame and the one obtained from TIM, when the target is kept in tracking. In Figure 1(b), there is a long distance between the target center of last frame and the one got from TIM, when the target is lost track.

2.2. Theoretical Justification of TIM

The proposed TIM approach is supported by the theoretical framework of Stable Marriage (SM) problem. The stable marriage problem is first introduced by Gale and Shapley in [6]. It is a matching problem between two sets of elements which are respectively named men set and women set by the initiators. Each man has their own preference list of the women they want to match (or marry), so does each woman. The task is to determine an assignment where each man is matched (or married) to one and only one women (monogamous and heterosexual). An instance of the preference lists are shown in Figure 2.

Men's List	Women's List
A: a, c, b, d	a: A, B, D, C
B: c, b, d, a	b: D, B, C, A
C: a, b, c, d	c: A, C, D, B
D: d, c, b, a	d: B, D, A, C

Figure 2. Sample preference list for men and women.

In Figure 2, each man, denoted by the list(A,B,C,...), has a list of women (a,b,c,...) ordered by his preference. Each woman has a similarity ranked list.

In the SM problem, a kind of dissatisfied pair is defined: when given a married pair, X-a and Y-b, if man X prefers woman b more than his current wife a and woman b prefers X more than her current husband Y, then X-b is called a dissatisfied pair. If there are no dissatisfied pairs, the assignment is said to meet the stable marriage criterion. The stable marriage assignment of Figure 2 is shown in Figure 3. The stability of this assignment can be verified according to the definition of dissatisfied pair.

Men's List	Women's List
A: a, c, b, d	a: A, B, D, C
B: c, b, d, a	b: D, B, C, A
C: a, b, c, d	c: A, C, D, B
D: d, c, b, a	d: B, D, A, C

Figure 3. Sample stable marriage for men and women.

To consider the target tracking problem in SM framework, the region of background is segmented into several virtual targets. The track loss is comprehended as a wrong match between the real target and one of the virtual targets. Under this framework, a single object tracking problem will be convert to a multi-object matching problem. However, it is not a multi-object matching problem as usual, because we only concentrate on the matching result of real target. Two preference lists of virtual and real targets can be obtained in the last and the current frame. The lists are ordered by similarity between these targets. Because the current target is the tracking result in last frame, the best match of last target is of course the current one. For this reason, the first element in the preference list of last target is current target for sure. For the same reason, if the result of TIM algorithm is consistent with the last target, the first element in preference list of current target is also the last target to a certainty. Under this situation, the last and the current targets are a stable marriage pair which can be guaranteed by the following theorem.

THEOREM 1. *If man A and woman a are both the best choice of each other in the preference lists, the married pair A-a is a stable marriage pair.*

PROOF. Using the male optimal stable solution proposed by Gale and Shapley in [6] to find a set of stable marriages, the man A will certainly propose to his first choice woman a in the first stage. If there is no other man also proposes to the woman a in all of the stages, then the married pair A-a will never be broken and will be kept as a stable marriage pair at last, else it will lie on the preference list of woman a. Assume that there is a man B who also proposes to the woman a in existence. According to the rule of the male optimal stable solution, the pair A-a which is preferred by woman a will be kept, and the pair B-a will be broken. Because A is the favorite man of woman a, the pair A-a can not be broken by any other man until the last stage. So the married pair A-a is certainly a stable marriage pair.

In visual tracking, targets pair which meets the stable marriage criterion is not certain a stable match. Because the set of stable marriages is not alone, two or three women may be the stable married wife of the same man in different stable marriage solutions. This can't be tolerated in object tracking, because multiple possibility of targets matching will lead to a unstable tracking result. To guarantee the stability of targets matching, the targets pair is not only necessary to be stable marriage pair, but also to be the exclusive stable marriage pair existing in all the possible set of stable marriages. This can be supported by the following THEOREM 2.

THEOREM 2. *If man A and woman a are both the best choice of each other in the preference lists, the married pair A-a is an exclusive stable marriage pair existing in all the possible sets of stable marriages.*

PROOF. The THEOREM 2 in [12] said that in any stable marriage no woman receives a poorer choice than the one she receives in the male optimal solution. In THEOREM 1 of this paper, the married pair A-a is proved to be a stable marriage pair in the male optimal solution. Since man A is the best choice of woman a, there is only poorer choices for the woman a. So woman a can not make another choice in all the sets of stable marriages, and the married pair A-a will be kept exclusively.

Thus if the target position of last frame is consistent with the result of TIM from current frame, the targets pair is a stable match, and the tracking result is creditable. Otherwise, the stability of the targets pair can not be guaranteed, and the track loss possibly occurs.

3. Improved Tracking System with Online Determination of Track Loss

Online determination of track loss can make improvements to the robustness of visual tracking systems. It mainly plays an important role in two aspects: 1) When track loss occur, trigger the template rematch in current frame. It can obtain a more reliable tracking result and do its best to avoid the system failure; 2) When keep track, collect the feature templates to save as standard ones for the template rematch in 1). But it is a problem to continuously collect the standard templates. Since the quantity of standard templates keep increasing, the template rematch will be time consuming. So an online templates clustering method is necessary to keep the quantity of standard templates.

The online templates clustering is a well studied problem in computer vision, and there are many methods can be used in the system [1] [2]. In this paper, we simply use the color histogram as the template feature, and adopt an online template clustering method similar to online k-means clustering [11]. The method to get standard templates based on online template clustering is shown in Table 1.

In Table 1, the quantity of standard templates for clustering, can be confirmed in advance according to the actual conditions. Instance of the head tracking using color feature, the quantity of standard templates must at least be two for the two kinds of head color, skin and hair. In the line 8 and 9 of Table 1, a weight w_k is introduced. It is got from the position match distance of TIM in the kth frame. The expression is shown as below:

$$w_k = \frac{1}{1 + d_k} \tag{4}$$

where d_k is the match distance of TIM which is defined in Expression 3. In Expression 4, if it is a perfect match between the last target and the result of TIM, the d_k equals zero, and the max value 1 of w_k can be reached. with the d_k increasing, the w_k in Expression 4 will reduce.

1: Initialize the standard templates number n and set the frame counter k to zero

2: LOOP

3: $k = k + 1$

4: IF $k < n + 1$ THEN

5: standard templates initialization: directly set the input template as one of the standard templates

6: ELSE

7: find the closest standard template to the input template in feature space

8: **compute the distance between the last target and the result of TIM to take as a weight** w_k

9: **standard templates update: online update the standard template found in line 7 using input template with the weight** w_k

10: END IF

11: END LOOP

Table 1. Method to get standard templates

The steps of taking advantage of track loss determining to improve tracking system can be summarized as follows:

1. Collect the standard templates when the target keep in track.

2. Online cluster the standard templates to keep the quantity of them.

3. Do a rematch by using the standard templates to obtain a more reliable result when the track loss occurs.

The diagram of the improved system based on track loss determining is shown in Figure 4.

In Figure 4, the determination of track loss is performed in each frame. The tracking result without track loss will be used to online update the standard templates. When the track loss occur, the standard templates are utilized for template rematch. The rematch result is also be determined whether the track loss take place.

4. Experimental Results

To show the performance of proposed TIM method, two parts of experiments are done: 1) An experimental comparison of track loss determination accuracy among TIM, WCD and SCD; 2) performance improvement for tracking systems using TIM.

The experiments are implemented on a standard PC (Pentium IV at 2.0GHz). The video image size is 320x240

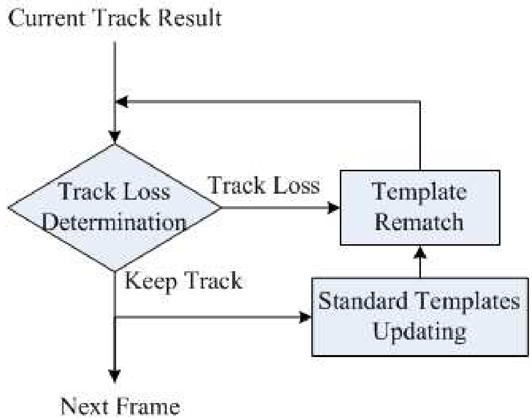


Figure 4. Diagram of the improved tracking system.

(24 bits per pixel) captured by surveillance cameras at 25fps. Mean Shift is used as the tracking algorithm framework, and multiple template features, such as color and gradient is utilized in our experiments. The tracking system using proposed track loss determination approach works at 15fps.

4.1. Experimental Comparison of Determination Accuracy

In this subsection, the proposed TIM method is compared with WCD and SCD which are introduced in section 1. The reason that we select the WCD and SCD to make a comparison is that they are also adaptive to all the visual tracking frameworks, and also can be done with low computational complexity which is necessary for visual tracking systems.

80 video frame sequences are captured as the experimental data. To keep the diversity and universality of the data, we capture them for some different objects, and in some different environments. Some sample frames of these data are shown in Figure 5.

There are four steps to compare the determination accuracy in this experiment. It is shown as follows:

1. For the video sequences in which the target is tracked successfully, record the determination values of track loss for the three methods in each frame, i.e. the similarity value of track matching for WCD, the covariance of similarity for SCD and the distance of inverse matching for proposed TIM.

2. For the video sequences in which the target lose track, record the determination values of track loss for the three methods in track loss occurrent frame.

Figure 5. sample frames of experimental data. The red rectangles show the target positions, and the targets from left frame to right frame are respective face, cup and phone

3. Draw two classes distributions. The values of negative samples are got from the first step, and the values of positive samples are got from the second step.
4. Move the determination threshold in the above two classes distributions to get the ROC curves.

The ROC curves of the three track loss determination methods in this experiment are shown in Figure 6.

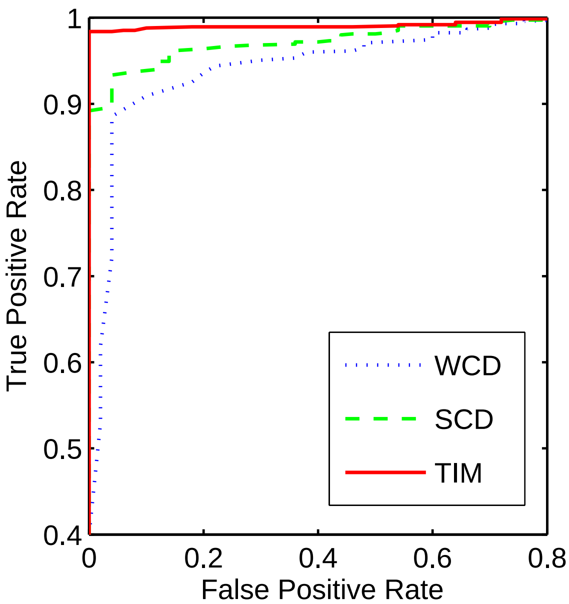


Figure 6. ROC curves of three different methods. ROC curves of TIM (red), SCD (green) and WCD (blue).

In Figure 6, It is shown that the performance of SCD is better than the performance of WCD. The TIM has the best performance whose accuracy rate is up to 98 percent. So the proposed TIM method can achieve high accuracy to determine track loss.

4.2. Experiments for Tracking Performance Improving

Using the approach presented in section 3, the performance of tracking systems can be made some improvement. An improved face tracking is shown in Figure 7 and an improved cup tracking is shown in in Figure 8. The proposed method can be also used in active tracking system, and this is shown in Figure 9.

In Figure 7, the face tracking without track loss determination is failed in frame 4, whereas the one with track loss determination keeps the target in track.

In Figure 8, the cup track is recovered by the track loss determination in frame 83.

In Figure 9, improved performance in active tracking system is shown. Using proposed method, the system keeps tracking in frame 83, whereas the tracking system without track loss determination is failed in this frame.

Through above experiments, it is verified that performance of tracking system can be improved by the proposed track loss determination strategy.

5. Conclusions

This paper presents a novel method to online determine track loss. It is automatic, without use of any ground truth, real time, adaptive to all the visual tracking frameworks and also with a high determination accuracy. The idea of the proposed method which inverse the common process of template tracking match is novelty. It is justified by the theoretical framework of Stable Marriage problem. With online determining track loss, the performance of tracking systems can be improved. The proposed method is proved to be practical and effective.

6. Acknowledgements

This work was performed in Biometrics and Security Research & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. It is supported by the following funds: Chinese National Natural Science Foundation Project #60518002, Chinese National 863 Program Projects #2006AA01Z192, #2006AA01Z193, and #2006AA780201-4, Chinese National Science and Technology Support Platform Project #2006BAK08B06, and Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

[1] H. A. Boubacar and S. Lecoeuche1. A new kernel-based algorithm for online clustering. In *Artificial Neural Networks:*

Formal Models and Their Applications - ICANN 2005, volume 3697, pages 583–588. Springer, 2005.

[2] A. Bouguettaya. On-line clustering. *Knowledge and Data Engineering*, 8:333–339 Vol.8, Apr 1996.

[3] C. Erdem, A. Murat Tekalp, and B. Sankur. Metrics for performance evaluation of video object segmentation and tracking without ground-truth. *In Proceedings of the International Conference on Image Processing*, Vol.2:69–72, Oct 2001.

[4] C. Erdem, B. Sankur, and A. Tekalp. Non-rigid object tracking using performance evaluation measures as feedback. *In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2:II–323–II–330, 2001.

[5] M. Franklin, M. Gondree, and P. Mohassel. Improved efficiency for private stable matching. In *In The Cryptographer's Track at RSA Conference (CT-RSA)*, pages 163–177. Springer, 2007.

[6] D. Gale and L. S. Shapley. College admissions and the stability of marriage. In *The American Mathematical Monthly*, volume 69, pages 9–15. Mathematical Association of America, 1962.

[7] I. K. M. S. Halldrsson, M. M. and H. Yanagisawa. Improved approximation results for the stable marriage problem. In *ACM Transactions on Algorithms (TALG)*, volume 3, page 30. ACM, 2007.

[8] H. Inoue, T. Tachikawa, and M. Inaba. Robot vision system with a correlation chip for real-time tracking,optical flow and depth map generation. *In Proceedings of the IEEE International Conference on Robotics and Automation*, Vol.2:1621–1626, May 1992.

[9] S. Y. H. Iwama, K. Miyazaki. Approximation algorithms for the sex-equal stable marriage problem. In *LECTURE NOTES IN COMPUTER SCIENCE*, volume 4619, pages 201–213. SPRINGER-VERLAG, 2007.

[10] L. Lu, X. tian Dai, and G. Hager. A particle filter without dynamics for robust 3d face tracking. In *in Proc. of CVPR Workshop on Face Processing in Video (FPIV04), Washington DC*, 2004.

[11] J. MacQueen. On convergence of k-means and partitions with minimum average variance. *Ann. Math. Statist.*, 36:1084, 1965.

[12] D. McVitie and L. Wilson. The stable marriage problem. In *Communications of the ACM*, volume 14, pages 486–490. ACM, 1971.

[13] L. Pao and W. Khawsuk. Determining track loss without truth information for distributedtarget tracking applications. *In Proceedings of the American Control Conference*, Vol.6:4363–4367, 2000.

[14] R. Powers and L. Pao. Detecting track-loss for the probabilistic data association filter in the absence of truth data. *The 43rd IEEE Conference on Decision and Control, CDC2004*, Vol.4:4316–4323, Dec. 2004.

[15] H. Wu, A. Sankaranarayanan, and R. Chellappa. In situ evaluation of tracking algorithms using time reversed chains. *In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, CVPR07.*, pages 1–8, June 2007.

[16] H. Wu and Q. Zheng. Self-evaluation for visual tracking systems. *In Proceedings of the 24th Army Science Conference*, Dec 2004.

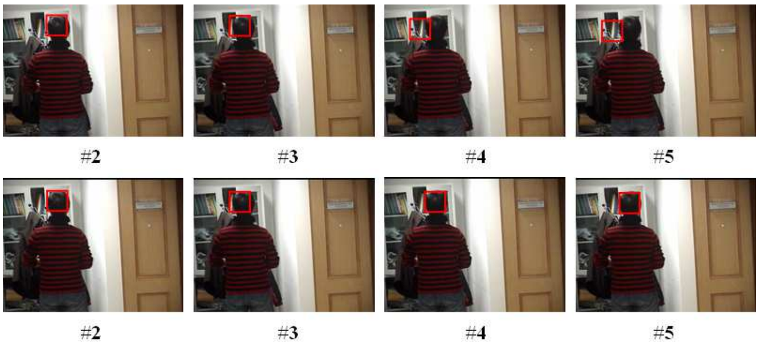


Figure 7. Improved performance in face tracking. The red rectangles show the estimated target positions by the tracking system. It shows the tracking result of video sequences from frame 2 to frame 4 without track loss determination (upper column) and with track loss determination (lower column).

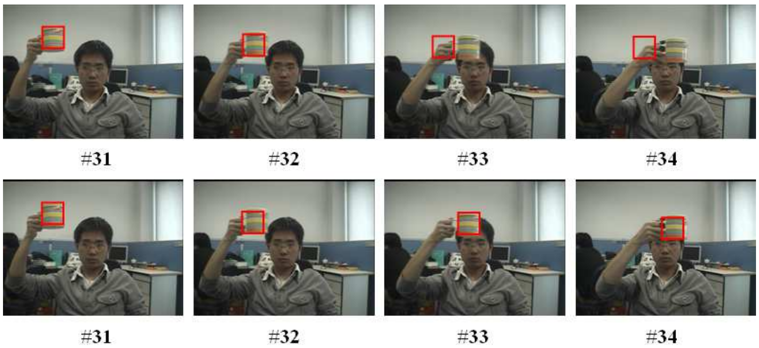


Figure 8. Improved performance in cup tracking. The red rectangles show the estimated target positions by the tracking system. It shows the tracking result of video sequences from frame 31 to frame 34 without track loss determination (upper column) and with track loss determination (lower column).

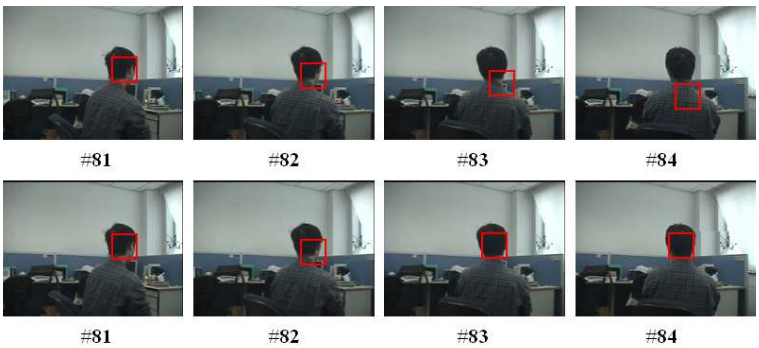


Figure 9. Improved performance in face active tracking. The red rectangles show the estimated target positions by the tracking system. It shows the tracking result of video sequences from frame 81 to frame 84 without track loss determination (upper column) and with track loss determination (lower column).