SCALE AND ROTATION INVARIANT FEATURE-BASED OBJECT TRACKING VIA MODIFIED ON-LINE BOOSTING

Quan Miao[1], *Guijin Wang*[1], *Xinggang Lin*[1], *Yongming Wang*[2], *Chenbo Shi*[1], *Chao Liao*[1]

[1]Tsinghua University, Department of Electronic Engineering, Beijing
[2] Advanced Information Technology Institute, Beijing

ABSTRACT

Object tracking is a major technique in image processing and computer vision. In this paper, we propose a new robust feature-based tracking scheme by employing adaptive classifiers to match the detected keypoints in consecutive frames. The novelty of this paper is that the design of online boosting is combined with the invariance of local features so that the classifier-based descriptions are formed in association with the scale and rotation information. Furthermore, we introduce a sample weighting mechanism in the on-line classifier updating, for the subsequent tracking. Experimental results demonstrate the robustness and accuracy of our proposed technique.

Index Terms— object tracking, keypoint matching, online boosting, classifier updating

1. INTRODUCTION

Robust object tracking is still a challenging task due to various variations like shape changes and appearance changes. Recently, tracking formulated as a classification problem has received a lot of attention because of its promising results. The classification-based tracking algorithm can be classified into two categories: region-based and feature-based method.

In case of the region-based methods [1, 2, 3], the basic idea is to learn a binary classifier which distinguishes the object from the background. However, these approaches have problems with complex transformations of the target object. In contrast, the feature-based trackers [4, 5] are more robust. In [4], a tracker employs randomized trees and ferns to discriminate keypoints from each other by classifiers. The disadvantage is that a considerable amount of time is spent in the off-line training phase. Also, tracking will fail if a certain appearance change of the object is not covered in the training phase.

To cope with these problems, Grabner proposes an efficient tracking approach [5] which employs the on-line boosting algorithm [6]. However, the keypoints are detected using Harris corner that is sensitive to scale changes. The tracker will fail when significant appearance variations such as affine transformation and viewpoint change arise.

This paper proposes a novel feature-based object tracking algorithm via online boosting. The main contributions are twofold. First, we utilize the robustness of local feature along with the adaptability of on-line boosting framework. The invariance of local feature to scale and rotation changes is fused in the learning of the classifier-based descriptions. This allows the tracker to handle significant transformations between frames. Second, we propose an adaptive scheme for classifier updating by assigning discriminative samples high importance weights. In comparison with [5], our tracker shows better performance on the challenging video sequences.

2. PROBLEM FORMULATION

The keypoint matching problem is formulated as follows. We build a P-class discriminant by constructing P classifiers $\{C_1, C_2, \ldots, C_P\}$ each corresponding to a keypoint $\{k_1, k_2, \ldots, k_P\}$ lying within the current object region. Given the keypoints set $\Upsilon = \{\gamma_1, \gamma_2, \ldots, \gamma_Q\}$ detected in the new frame, we employ the classifiers to find the point ε_i corresponding to k_i by:

$$\varepsilon_i = \underset{\gamma_q \in \Upsilon}{\arg\max}\, C_i(\gamma_q), \tag{1}$$

Similarly, the set of matching candidates $\Sigma = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_P\}$ is established by evaluating all the classifiers on Υ.

In the following, we briefly review the online boosting algorithm [6] in which the boosted classifier C is composed of J selectors h_j^{sel}. Fig.1 gives the framework of online boosting. Each classifier holds a weak classifier pool X from which the training procedure selects the ones with the minimal estimated error. The classifier wishes to predict the matching confidence measure of an unknown point $\mathbf{x}$ by:

$$C(\mathbf{x}) = conf(\mathbf{x}), \tag{2}$$

$$conf(\mathbf{x}) = \sum_{j=1}^{J} \alpha_j \cdot h_j^{sel}(\mathbf{x}) \Bigg/ \sum_{j=1}^{J} \alpha_j, \tag{3}$$

where $conf(\bullet)$ denotes the confidence measure. $h_j^{sel}(\mathbf{x}) \in \{1, -1\}$ predicates the label of $\mathbf{x}$. As new samples arrive sequentially, each selector h_j^{sel} is responsible for selecting the

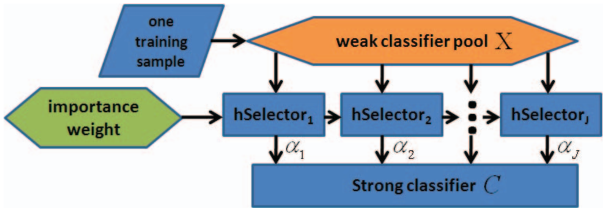

Fig. 1. On-line boosting for feature selection.

best weak classifier and the voting weight α_j is updated. This update is done with respect to the importance weight of the current sample.

During boosting learning, how to construct the robust weak classifier pool is an important issue. The method described in [6] uses the standard Haar-like features [7] that are computed in a fixed bounding patch centered at the corresponding keypoint. Nevertheless, these simple features can only deal with invariance under pure translations and slight rotations. We improve this scheme by incorporating the scale and the dominant orientation of the keypoint in the weak classifier pool. In addition, each sample should bear an importance weight to indicate its contribution to the classifier updating. Grabner's method gives all the samples equal weight. This paper proposes to emphasize the negative samples that are "similar" to the positive one, to make the updated classifier more discriminative.

3. PROPOSED ALGORITHM

The region of the target object is defined in the first frame. When the new frame arrives, we first detect the keypoints. Then we compute the classifier-based descriptors and perform keypoint matching with the previous frame. The homography H is estimated using RANSAC [8] over the set of matching candidates. If the number of inliers is below a threshold, we apply no updates to the classifiers and skip to the next frame. Otherwise, tracking is considered successful by geometric transformation using H. In addition, the object representation is updated to perform target tracking in the subsequent frame. In the remainder of this section we will describe the algorithm shown in Fig.2.

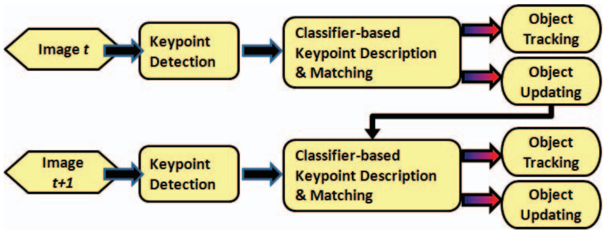

Fig. 2. Flowchart of the proposed algorithm.

3.1. Keypoint detection

For keypoint detection [9], we choose the SURF feature detector [10] because it has outperformed most of the previous schemes. Meanwhile, it is fast due to the use of integral images.

3.2. Scale and rotation invariant classifiers for keypoints

In this subsection, we integrate the scale and the dominant orientation information of each detected keypoint in the weak classifiers to make the descriptions invariant to many kinds of geometric transformations. The scale s is obtained simultaneously with the location detecting. The corresponding dominant orientation α is estimated within a circular neighborhood centered at the keypoint; the radius is proportional to the scale.

Similar to [6], this paper also relates each weak classifier with a Haar wavelet representation to capture the structural similarities between object changes. The difference is that our neighborhood used to compute the Haar-like responses corresponds to a square region centered around the keypoint and oriented along α. The size of this window is $\mathrm{M}s$ (M is a constant). Our approach samples this window with a sampling step to be s, and then computes the Haar wavelet responses at each sample point. Furthermore, the responses are distributed in horizontal and vertical direction where "horizontal" and "vertical" is defined in relation to α, to achieve rotation invariance. Finally, scale invariance is realized by normalizing each response with respect to $s \times s$. Fig.3 shows the construction of the weak classifier pool X.

As can be seen, our classifier-based keypoint description is quite different from the SURF feature description despite the same keypoint detection. As for the SURF descriptor, spatial distribution of gradient information has to be focused on and an interpolation of the histogram should be added to form the high-dimensional vectors. In contrast, our approach of using Haar wavelet responses as the weak classifiers is simpler and less computationally intensive.

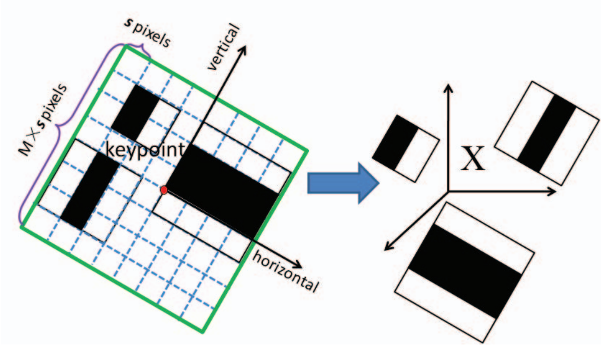

Fig. 3. The construction of a weak classifier pool.

3.3. Object tracking

Assuming the target object has been successfully identified in frame $t-1$, we use RANSAC to estimate the homography $H_{t,t-1}$ in frame t. If the number of inliers exceeds a threshold, the object can be tracked by transforming the object region in frame $t-1$ using $H_{t,t-1}$.

3.4. Object updating

If tracking is successful, the classifiers describing the tracked object will be updated by taking matches as positive samples and other keypoints as negative. In our approach, the importance weight of each negative sample $\mathbf{x}$ consists of two parts. First, it is connected with its confidence measure:

$$\lambda_1(\mathbf{x}) = \mu + \sigma \cdot \exp\{conf(\mathbf{x}) + \eta\}, \qquad (4)$$

where μ, σ and η are constants. When the confidence measure for a sample is high, the weight is likewise high for that sample. This means that the online trained classifier is less likely to have false positives, as shown in Fig.4. As a result, the matching accuracy is ensured by the updated classifiers.

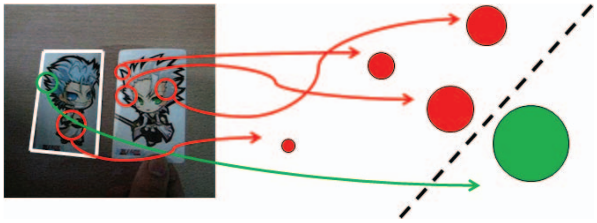

Fig. 4. Negative samples that are closer to the classifier hyperplane are assigned higher importance weight.

Furthermore, as RANSAC is used in motion estimation, false matching candidate is prone to be located near the true correspondence in the image domain. Thus we introduce the kernel function, to emphasize the nearer negative samples:

$$\lambda(\mathbf{x}) = \lambda_1(\mathbf{x}) \cdot K(\mathbf{x} - \hat{\mathbf{x}}), \qquad (5)$$

where $K()$ is the 2-D realization of the kernel function, which is symmetric and attains its maximum at zero. $\mathbf{x}$ and $\hat{\mathbf{x}}$ respectively denote the selected negative sample and the positive sample. By this means, each classifier can better discriminate the corresponding keypoint from its possible confused neighbor. The importance weight of the positive sample $\hat{\mathbf{x}}$ is the summation of the weights of all the negative samples.

4. EXPERIMENTAL RESULTS

We now present the experimental results of applying our algorithm on several video sequences. For comparison, we implemented Grabner's tracker [5]. To better illustrate the soundness of our approach, we have implemented another tracking method in which the first frame is considered as the reference frame and correspondences are established between the keypoints in the defined object region of the reference frame and those in the input frame. The best candidate match is defined as the one with the minimum Euclidean distance for the SURF descriptor vector. We call this second approach the SURF-based method. In experiment section, we will compare the performance of the SURF-based method, Grabner's method as well as the proposed method.

4.1. Object tracking under viewpoint change

As a first attempt, we focus on tracking a book cover in the following sequence in which we open and close the book cover to generate a viewpoint change. Fig.5 shows the tracking performance. It can be observed that the SURF-based method fails to track the object in later frames, since it lacks the adaption to the motion of the object. For Grabner's approach, tracking fails once the viewpoint change becomes severe because the Harris corner detector cannot find relevant keypoints. Although it re-tracks the object afterwards, the identified region is heavily distorted. As for the proposed method, the combination of SURF detector's high detection accuracy and the on-line adaption to the changes of the object helps to preserve satisfying tracking despite the significant appearance change.

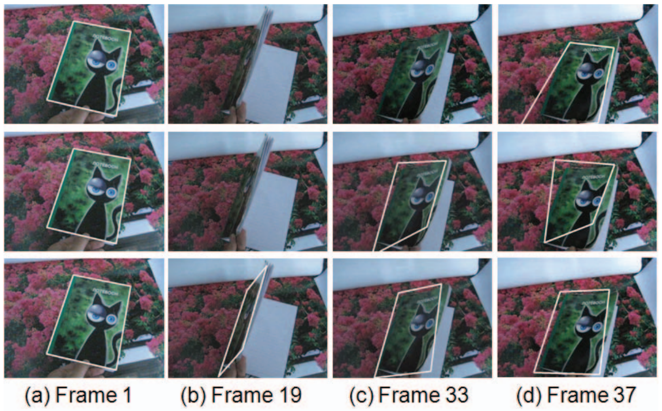

Fig. 5. Tracking a book cover under viewpoint change. From top to bottom row, tracking result using SURF-based tracker, Grabner's tracker and the proposed tracker.

Fig.6 shows the number of matches certificated by RANSAC for each frame. Tracking loss (the percentage is below 25%) occurs continuously using SURF-based method and Grabner's method (from frame 11 to frame 31) because of significant appearance change. However, our proposed tracker can handle the phenomenon, except around frame 22 where the object drastically changes its appearance so that few keypoints are detected. As for most of the frames with successful tracking, the classifiers in the proposed algorithm establish

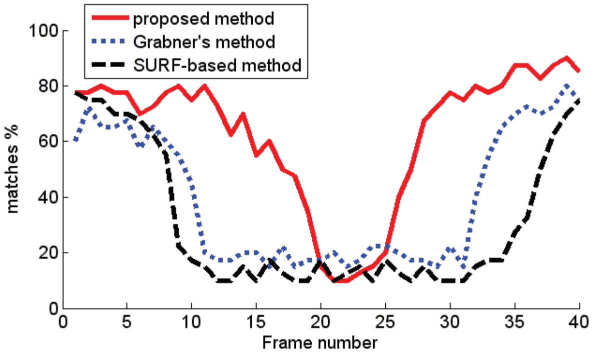

Fig. 6. Number of the matches of the proposed algorithm versus the SURF-based method and Grabner's method.

more correct matches than those in the other two methods, validating the superiority of our approach.

4.2. Moving target with scale and rotation variation

Furthermore, rapid scale and rotation variations between frames do not confuse our tracker, as is shown in Fig.7. The sequence is captured by moving the target up and down and rotating it. Grabner's tracker fails under rapid changes, since these changes cannot be handled only by exchanging features. The SURF-based tracker could work in most cases but not all; sometimes the appearance change between the current frame and the first frame is too significant for the SURF descriptor to preserve invariant. In contrast, our work advances the SURF-based method by providing the robust formula of on-line classifier-based keypoint matching. The classifiers' invariance to scale and rotation and their discriminative power make the proposed method ideally suited for this kind of issue.

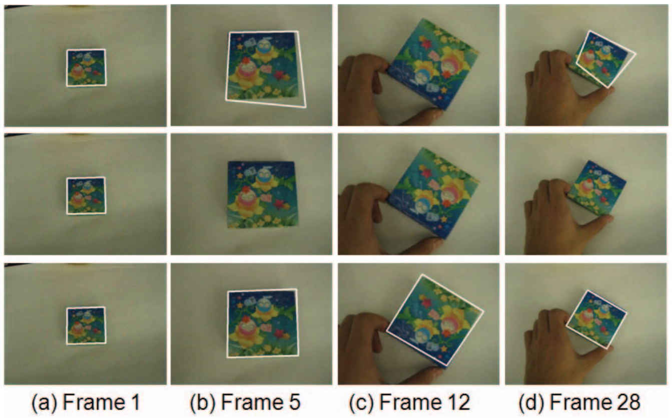

Fig. 7. Tracking a notebook under rapid scale and rotation change. From top to bottom row, tracking result using SURF-based tracker, Grabner's tracker and the proposed tracker.

5. CONCLUSION

This paper presents a new feature-based technique treating object tracking as a keypoint matching problem. The proposed approach demonstrates that incorporating the robust local feature and the adaptive online boosting algorithm can efficiently cater to changes between successive frames. Moreover, the modified object updating scheme improves the discriminative power of the classifiers. Experimental results verify that our approach completely outperforms the state-of-art tracker to achieve robust and accurate object tracking.

6. REFERENCES

[1] S. Avidan, "Ensemble tracking," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 494–501.

[2] R. T. Collins, Yanxi Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1631–1643, Oct. 2005.

[3] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. European Conference on Computer Vision*, 2008.

[4] V. Lepetit, P. Lagger, and P. Fua, "Randomized trees for real-time keypoint recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 775–781.

[5] M. Grabner, H. Grabner, and H. Bischof, "Learning features for tracking," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[6] H. Grabner and H. Bischof, "On-line boosting and vision," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[8] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[9] J. Li and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, no. 10-12, pp. 1771–1787, 2008.

[10] H. Bay and A. et al Ess, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.