The Effect of 2nd-Order Shape Prediction on Tracking Non-Rigid Objects

Kenji Nishida
and Takumi Kobayashi

National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, JAPAN
Email: kenji.nishida@aist.go.jp takumi.kobayashi@aist.go.jp

Jun Fujiki

Department of Applied Mathematics,
Fukuoka University
Fukuoka, JAPAN
Email: fujiki@fukuoka-u.ac.jp

***Abstract*—**

For the tracking of non-rigid objects, we have previously proposed a shape-based predict-and-track algorithm. The method was built upon the similarity between the predicted and actual object shapes. The object shape was predicted from the movement of feature points, which were approximated by a second-order Taylor expansion. Approximate first-order movements, the so-called optical flows, were simultaneously exploited by chamfer matching of edgelets. In this paper, the effect of second-order shape prediction is quantitatively analyzed by tracking a non-rigid skier object. The method exhibits superior shape prediction performance compared to a simple linear prediction method.

Keywords–Tracking non-rigid objects, Chamfer distance, Shape prediction.

I. INTRODUCTION

Visual object tracking is one of the most popular techniques in the field of computer vision. Recently, tracking algorithms for non-rigid (deformable) objects have been used in many application fields [1], [2]. In sports scenes, especially those of team sports such as football, there are many similar objects, which increase the difficulty of tracking. Therefore, we consider both the movement and form (shape) of these objects to be discriminative for tracking.

We have already proposed a shape-based predict-and-track algorithm [3], which was evaluated by tracking a skier to determine the effect of shape prediction performance. In this paper, we quantitatively evaluate the shape prediction performance of a second-order shape prediction algorithm against linear (first-order) prediction. The performance is measured by the similarity between the predicted and actual shapes of the tracked object.

The remainder of this paper is organized as follows. In Section II, we describe our shape prediction algorithm and the tracking procedure that uses the chamfer distance as a similarity measure. The experimental results are presented in Section III. Finally, we present our conclusions and ideas for future work in Section IV.

II. SHAPE-BASED PREDICT-AND-TRACK ALGORITHM

In this section, we describe an algorithm for tracking by shape prediction [3]. The algorithm consists of two components, shape prediction and tracking by shape similarity.

A. Notation

The following notation is used throughout this paper.

- X denotes the center of the object,
- $O(X)$ denotes the object image centered at position X,
- $E(X)$ denotes the binary edge image for the object at position X,
- $\hat{O}$ and $\hat{E}$ denote the predicted image and edge image of the object, respectively,
- x denotes the positions of the feature points for object X,
- x' denotes the differential of x, i.e., $x' = \frac{dx}{dt}$,
- x'' denotes $\frac{d^2x}{dt^2}$,
- $\tilde{x}$ denotes the subset of feature points in the object that constitute the outline edge, $\tilde{x} \in E(X)$,
- $\hat{x}$ denotes the predicted position at the next frame for $\tilde{x}$,
- $l(x)$ denotes the edgelet for position x.

B. Shape Prediction

The object shape is represented by the collection of feature points x, and the deformation of the object is predicted by exploiting the movement of the feature points. Sim and Sundaraj proposed a motion tracking algorithm using optical flow [4], and this can be considered as the first-order approximation of the movement. For our tracking algorithm, we adopt a shape prediction method based on the second-order approximation of the feature points' movement [3].

Let x_t be the 2-D position of the feature points that constitute the object image O at time t. The position of the points at $t+1$ can be estimated using a Taylor expansion. Up to the second-order, this is

$$x_{t+1} = x_t + x'_t + \frac{1}{2}x''_t, \tag{1}$$

where x' is the so-called optical flow, which is practically computed as the difference in the pixel position:

$$x'_t = x_t - x_{t-1}. \tag{2}$$

Similarly, x'' denotes the second-order differential of x, which is calculated as

$$\begin{aligned} x''_t &= x'_t - x'_{t-1} \\ &= x_t - x_{t-1} - (x_{t-1} - x_{t-2}) \\ &= x_t - 2x_{t-1} + x_{t-2}. \end{aligned} \tag{3}$$

 ISBN: 978-1-61208-393-3

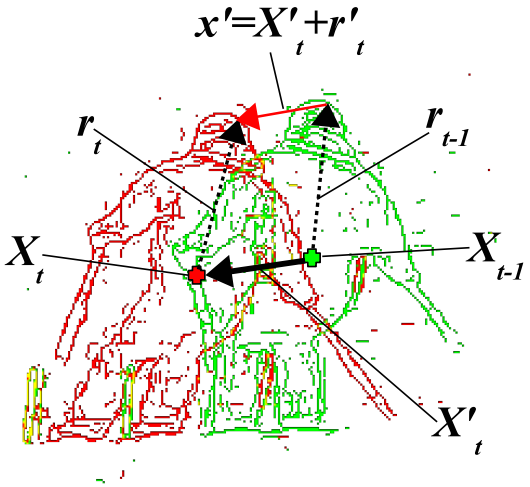
Therefore, the appearance of the object at $t+1$ can be predicted based on the optical flows computed from three consecutive video frames. Suppose that the shape of the object is determined by the outline edge image E, predicted from the feature point movements in previous video frames. The algorithm for detecting the feature point movements is described in Section II-D.

C. Estimation of Object Translation

The movement of the feature points comprises both the object translation (global movement of the center of the object) and the movement of the pixels relative to the center of the object, which is described by

$$x'_t = X'_t + r'_t, \tag{4}$$

where X denotes the position of the object's center, and r denotes the position of the pixels relative to X. Figure 1 shows the movement of feature point x', the movement of the object's center X', and the relative movement r'.



Green: Edge image for $t-1$, Red: Edge image for t

Figure 1. Edge image and object movement.

The relative movement r' is derived from the object deformation, and thus makes a significant contribution to the prediction of the object's shape. Because relative movement obeys the physical constraints of the body parts of the object, its second-order prediction is effective. In contrast, the second-order movement contributes less to the object translation X, because such global movement contains the ego-motion of the camera as well as the real movement of the object. Therefore, the purpose of our tracking algorithm is to determine the next object position X_{t+1} based on the similarity between the predicted and actual object shapes, which is computed globally.

The similarity between the predicted edge image $\hat{E}_{t+1}$ and actual edge image E_{t+1} is measured using the chamfer system [5]. This system measures the similarity of two edge images using a distance transform (DT) methodology [6].

Let us consider the problem of measuring the similarity between template edge image E_t (Figure 2(b)) and a successive edge image E_{t+1} (Figure 2(c)). We apply the DT to obtain an image D_{t+1} (Figure 2(d)), in which each pixel value d_{t+1}

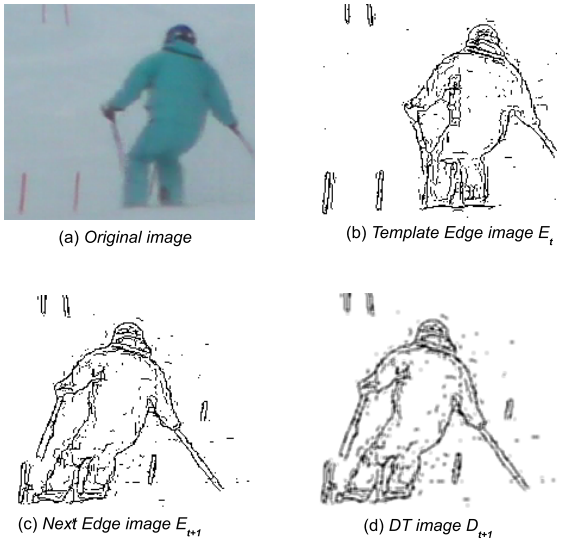


Figure 2. Chamfer system.

denotes the distance to the nearest feature pixel in E_{t+1}. The chamfer distance $D_{chamfer}$ is defined as

$$D_{chamfer}(E_t, E_{t+1}) = \frac{1}{|E_t|} \sum_{e \in E_t} d_{t+1}(e), \tag{5}$$

where $|E_t|$ denotes the number of feature points in E_t and e denotes a feature point of E_t.

The translation of the object can be estimated by finding the position of the predicted edge image $\hat{E}_{t+1}$ that minimizes $D_{chamfer}$ between $\hat{E}_{t+1}$ and the actual edge image E_{t+1}:

$$X_{t+1} = \arg\min_{\hat{E}_{t+1}} D_{chamfer}(\hat{E}_{t+1}, E_{t+1}). \tag{6}$$

Figure 3 illustrates the tracking procedure. First, the optical flow x'_t and its approximate derivative x''_t are computed from preceding video frames at $t-2$, $t-1$, and t. The object shape at $t+1$, denoted by $\hat{E}_{t+1}$, is then predicted using x' and x''. The object position is determined by locating $\hat{E}_{t+1}$ at the position of minimum chamfer distance to the actual shape at $t+1$, E_{t+1}. Finally, the optical flow for the next video frame x'_{t+1} is recomputed using actual edge images E_t and E_{t+1}.

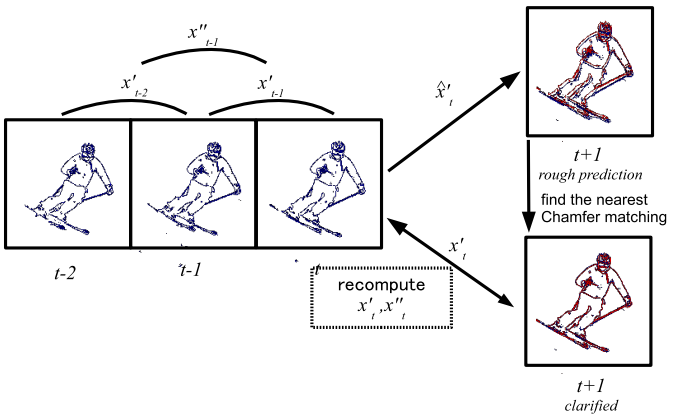


Figure 3. Tracking procedure.

D. Detection of Feature Point Movements

After the object translation X'_{t+1} has been determined, the movement of the feature points x'_{t+1} is detected from the actual object images $O(X_t)$ and $O(X_{t+1})$.

The feature point movements x'_{t+1} are directly computed based on the actual edge image at $t+1$ by tracking small parts of the edge (edgelets). We also employed the chamfer system to detect the movement of the edgelets. A template edgelet image $l(\tilde{x}_t)$ extracted from E_t is compared against the candidate edgelet $l(\tilde{x}_t + \hat{x}'_{t+1})$ in the next edge image E_{t+1}. By minimizing the chamfer distance between the two, we obtain the feature point movement (Figure 4):

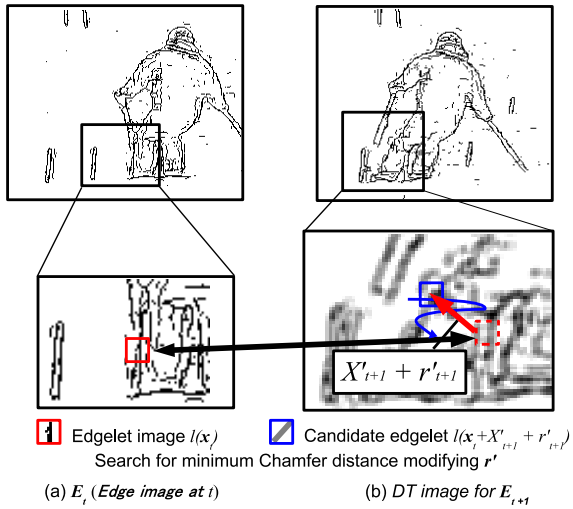


Figure 4. Edgelet tracker.

$$\hat{x}'_{t+1} = \arg\min_{\hat{x}'_{t+1}} D_{chamfer}(l(\tilde{x}_t), l(\tilde{x}_t + \hat{x}'_{t+1})). \quad (7)$$

As the detected movements $\hat{x}'_{t+1}$ may contain noise, we apply a smoothing process by averaging the relative movements in the neighboring region:

$$x'_{t+1} = \frac{1}{N} \sum_{\hat{x}'_{t+1} \in \delta_{t+1}} \hat{x}'_{t+1}, \quad (8)$$

where N denotes the number of detected movements $\hat{x}'_{t+1}$ in the neighborhood δ of $\tilde{x}_t$.

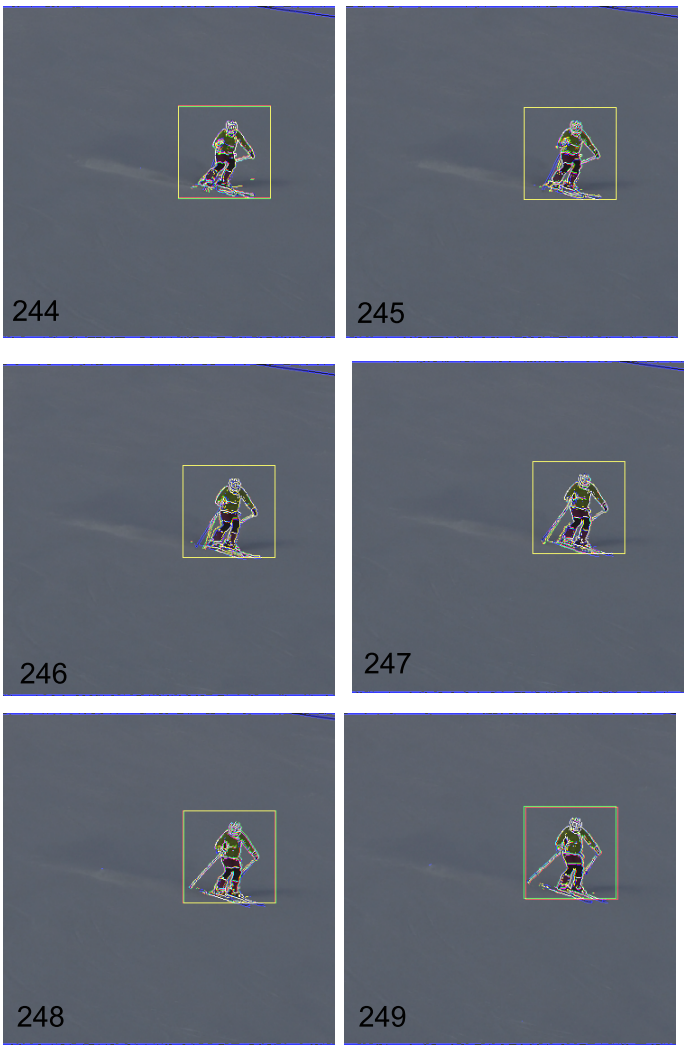
III. Evaluation of Shape Prediction Performance

The algorithm described above was applied to a video sequence of a skier, captured by a hand-held camera, and the effect of camera ego-motion on the shape prediction performance was examined. The second-order shape prediction was compared with linear shape prediction, which is formulated by

$$x_{t+1} = x_t + x'_t. \quad (9)$$

The prediction performance was evaluated by the chamfer distance between the actual image E_{t+1} and the predicted image $\hat{E}_{t+1}$.

In the skiing sequence captured by a hand-held camera, the skier was manually "tracked" so as to remain close to the center of the image frame. Thus, the object tends to exhibit only a small translation in the image frame. However, the object sometimes suffers from a large degree of translation due to manual mis-tracking of the camera. Figure 5 shows the tracking results. The blue pixels represents the predicted object shape, the green ones represent the translated predicted shape to determine the object position using (6), and the red ones represents the reconstructed object shape, as calculated by (1).



Blue: Ground Truth; Green: Linear Prediction;
Red: Second-order Prediction.

Figure 5. Tracking result.

Figure 6 shows the chamfer distance to the ground truth, calculated over frames 230–300. The results show that the second-order prediction attained better precision than the linear prediction in 40 out of 70 frames. The second-order prediction is superior during frames 244–249, whereas the linear prediction is preferable from frames 238–240.

Figure 7 shows the object translation from frames 244–248, indicating the direction change at around frame 246. Figure 8 shows the object translation from frames 238–240, when the translation direction did not change.

These results indicate that the second-order shape prediction method works well when the direction in which the object

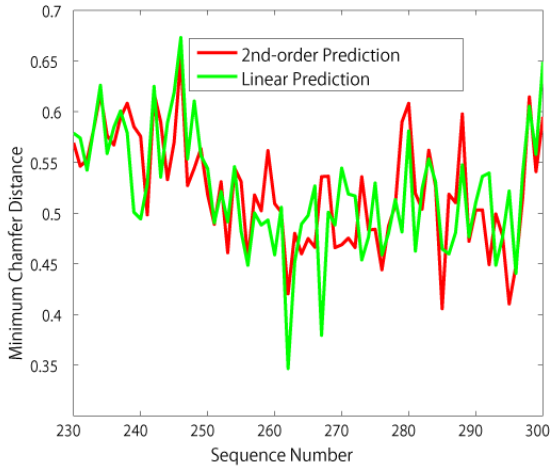
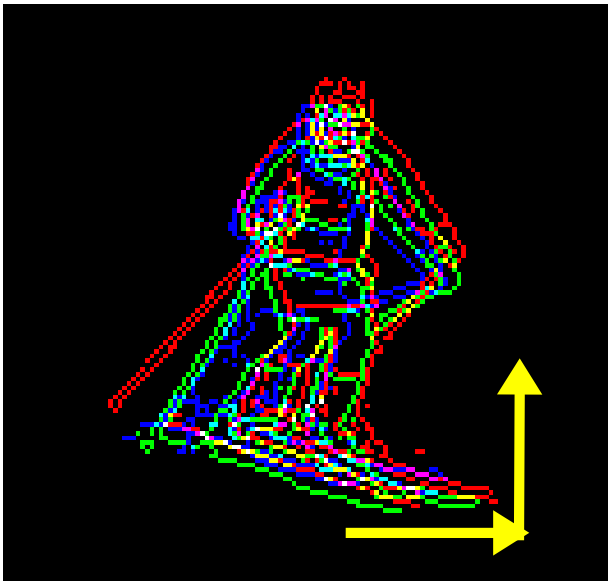


Figure 6. Shape precision according to chamfer distance.



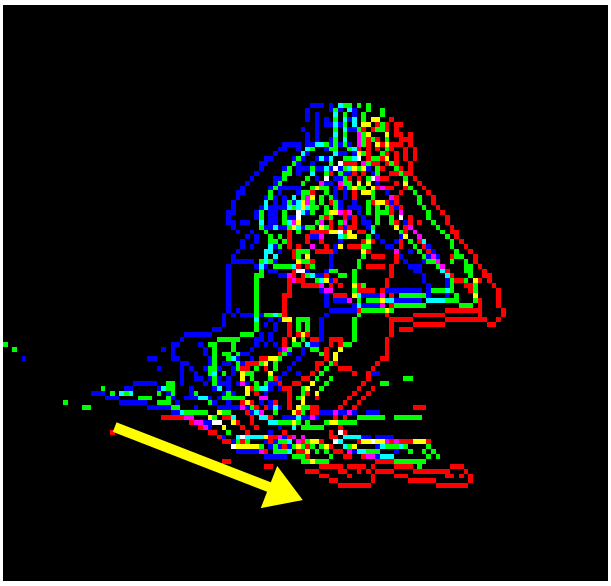
Blue: 244; Green: 246; Red: 248.
Yellow arrow: Object translation.

Figure 7. Object translation from frames 244–248.

must be translated changes.

IV. Conclusions

We have evaluated the performance of a second-order shape prediction algorithm. Though the performance is generally similar to that of a linear model, our method outperformed the linear approach when the direction of object movement changed. This evaluation result indicates that the proposed second-order model is robust to objects under high acceleration. In future work, we will integrate the two models, allowing us to switch which model is applied for prediction, and apply the method to various types of objects besides skiers.



Blue: 238; Green: 239; Red: 240.
Yellow arrow: Object translation.

Figure 8. Object translation from frames 238–240.

Acknowledgment

The authors would like to thank Dr. Akaho, group leader of Mathematical Neuroinformatics Group, for his valuable comments and suggestions. This work was supported by JSPS KAKENHI Grant Number 26330217.

References

[1] G.Sundaramoorthi, A.Mennucci, S.Soatto, and A.Yezzi, "A New Geometric Metric in the Space of Curves, and Applications to Tracking Deforming Objects by Prediction and Filtering", in SIAM j. of Imaging Science, Vol.4, No.1, 2010, pp.109-145.

[2] M. Godec, P.M Roth, and H.Bischof, "Hough-based Tracking on Non-rigid Objects", J. of Computer Vision and Image Understanding, Vol.117, No.10, 2013, pp.1245-1256.

[3] K. Nishida, T. Kobayashi, and J. Fujiki, "Tracking by Shape with Deforming Prediction for Non-Rigid Objects", in proc. International Conference on Pattern Recognition Applications and Methods (ICPRAM) , 2014, pp. 581-587.

[4] K.F. Sim, and K. Sundaraj,. "Human Motion Tracking of Athlete Using Optical Flow & Artificial Markers", in Proc. International Conference on Intelligent and Automation Systems (ICIAS), 2010, pp. 1-4.

[5] D.M. Gavrila, "Pedestrian Detection from a Moving Vehicle", in Proc. European Conference on Computer Vision (ECCV), 2009, pp. 37-49.

[6] D.Huttenlocher, G.Klanderman, and W.J.Rucklidge, "Comparing Images using the Hausdorff Distance", in IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No. 9, 1993, pp. 850-863.

 ISBN: 978-1-61208-393-3