Recognizing Static Target in Video Frames Taken from Moving Platform

Xin Wang*,**, Masanori Sugisaka**, and Wenli Xu*
*Department of Automation, Tsinghua University, Beijing, China
(Tel: 0081-90-8836-4099; E-mail: wang@oita-it.ac.jp, wlxu@tsinghua.edu.cn)
**Department of Electrical and Electronic Engineering, Oita University, Oita, Japan
(Tel: 0081-97-5547831; E-mail: msugi@cc.oita-u.ac.jp)

Abstract: This paper deals with the problem of moving object detection and location in computer vision. We describe a new object-dependent motion analysis method for tracking target in an image sequence taken from a moving platform. We tackle these tasks with three steps. First, we make an active contour model of a target in order to build some of low-energy points, which are called *kernels*. Then we detect interest points in two windows called *tracking windows* around a *kernel* respectively. At the third step, we decide the correspondence of those detected interest points between *tracking windows* by the probabilistic relaxation method In this algorithm, the detecting process is iterative and begins with the detection of all potential correspondence pair in consecutive image. Each pair of corresponding points is then iteratively recomputed to get a globally optimum set of pairwise correspondences.

Keywords: target, interest point, correspondence

1. INTRODUCTION

Motion detection is made difficult as both the observer and some elements of the scene may be moving. But detecting objects from a moving platform is one of the key issues to the successful applications of mobile system.

In recent year, interest in motion processing has increased with advances in motion analysis methodology and processing capabilities. The usual input to a motion analysis system is a temporal image sequence, with a corresponding increase in the amount of processed data. Motion analysis is often connected with real-time analysis, for example, for robot navigation. Another common motion analysis problem is to obtain comprehensive information about moving and static objects present in a scene. A set of assumption can help to solve motion analysis problem-prior knowledge helps to decrease the complexity of analysis. Prior knowledge includes information about the camera motion, mobile or static, and information about the time interval between consecutive images, especially whether this interval was short enough for the sequence to represent continuous motion. This prior information about data helps on the choice of an appropriate motion analysis technique.

There are three groups of motion-related problems from the practical point of view: motion detection, moving object detection and location and derivation of 3D object properties. We will deal with the second group, etc. moving object detection and location.

The notable algorithms of objective from a moving platform using a vision sensor camera use temporal or spatial disparities estimated from the image sequence for detection purpose. These algorithms can be grouped into three classes:

- methods using optical flow;
- methods using qualitive estimates of motion;
- methods using velocity field.

Object motion parameter can be derived from computed optical flow vectors [1][2]. But in reality, estimates of optical flow or point correspondence are noisy, three-dimensional interpretation of motion is ill-conditioned and requires high precision of optical flow or point correspondence.

To overcome this problem, motion field technique, which is based on images acquired at intervals that are not short enough to ensure small changes due to motion and can be also acquired if the number of images in a sequence is small, have appeared. Motion filed or velocity field computations represent a compromise technique; information similar to the optical flow is determined, but it is based on images acquired at intervals that are not short enough to ensure small changes due to motion. The velocity field can also be acquired if the number of image in a sequence is small.

Motion field analysis relies on distinctive characteristics in different frames and can be concluded into the point or area correspondence analysis. So we discuss motion analysis problem based on tracking correspondence parts in image sequence from a moving platform.

2. OVERVIEW

We present a new object-dependent algorithm that analyzes the behavior of the objects in an image sequence from a moving platform. We tackle these tasks with three broad approaches. First, we make a model of a target or object. This work is aimed at selecting some *kernels* about the target/object based on prior knowledge, limiting the target to a local area and also decreasing time cost. Another advantage of this selection is that we can estimate the object based on part of its *kernels*. We have augmented the general idea of snake [3]. The principle of detecting interest points around the *kernels* in an image frame are introduced in the second part. The third part of our strategy is a probabilistic relaxation method for detecting the correspondence of feature points. In this algorithm, the detecting process is iterative and begins with the detection of all potential correspondence pair in consecutive image. Each pair of corresponding points is then iteratively recomputed to get a globally optimum set of pairwise correspondences. Those feature points in consecutive image are used to initialize active contours as the input of next image frame. The part one and part two's works are repeated until reaching a new converged contour. An incomplete contour can also be established by prior knowledge. This is also a clue for us to solve the occlusion problems but not in this paper.

One of important differences between motion in moving camera or moving platform and static camera is that all is moving in the former case and only object are moving in the later. We discuss the case of a camera on a moving platform and the target/object is static on the assumptions that motion is with maximum velocity constraint and mutual correspondence that each point of an object corresponds to exactly one point in the next image in sequence in this paper.

Our approach consist of building a contour model of a target/object, and finding corresponding key areas or points in consecutive image which kernels appeared in the model. To get a corresponding relation, the first step is to decide those interest points and seek optimal correspondence or possible solutions. We denote $I(\mathbf{x})$, $\mathbf{x}$=(x, y) and $\mathbf{x} \in \mathcal{R}^2$, as any a frame in an image sequence and the i_{th} $(i = 1, 2, \ldots, n)$ frame is $I_i(\mathbf{x})$.

Fig. 1: An active contour model.

3. SNAKE MODEL

The traditional active contour model, snake, is defined as an energy minimizing spline–the snake energy depends on its shape and location within an image. The energy functional to be minimized may be written as:

$$E_s^* = \int_0^1 (E_{int}(\mathbf{v}) + E_{ext}(\mathbf{v}))ds. \tag{1}$$

where $\mathbf{v}=v(s)=[x(s), y(s)]$, x(s), y(s) are x, y co-ordinates along the contour and $s \in [0,\ 1]$, $v(s)$ can be approximated as a B-spline. $E_{int}(\mathbf{v})$ represents the internal energy of the spline due to bending. External energy $E_{ext}(\mathbf{v})$ includes image forces $E_{im}(\mathbf{v})$ and external constraint forces $E_{cst}(\mathbf{v})$:

$$E_{ext}(\mathbf{v}) = E_{im}(\mathbf{v}) + E_{cst}(\mathbf{v}). \tag{2}$$

$E_{im}(\mathbf{v})$ is derived from the image data over which the snake lies. $E_{cst}(\mathbf{v})$ come from external constraints which may force the snake toward or away from particular features.

The functional to be minimized is (1). This minimizing condition can reduce to

$$\frac{\partial v(s,t)}{\partial t} - \frac{\partial}{\partial s}[\alpha(s)\frac{\partial v(s,t)}{\partial s}] + \frac{\partial^2}{\partial s^2}[\beta(s)\,\frac{\partial^2 v(s,t)}{\partial s^2}] - \bigtriangledown E_{ext}[v(s,t)] = 0. \tag{3}$$

where $\alpha(s)$, $\beta(s)$ specify the elasticity and stiffness of the snake. We given a snake contour with N nodes, n_i=(u_i, w_i), $i = 1, 2, \ldots, N$, the state of the snake is represented by ($\mathbf{u}$, $\mathbf{w}$)=((u_1, ..., u_N)^T, (w_1, ..., w_N)^T). The resolution above the Euler-Lagrange equation (3) can then be written as:

$$\mathbf{AZ} + \mathbf{\Sigma F} = -\gamma \mathbf{\Delta Z}. \tag{4}$$

where $\mathbf{A}=\mathbf{A}(\alpha, \beta)$ is a pentadiagonal banded matrix that depends on α and β which control the internal spline energy of the snake. $\mathbf{Z}=\mathbf{Z}(\mathbf{u}, \mathbf{w})$, $\mathbf{F}$ denotes image external force. We define $\mathbf{F}_{im}=E_{im}(\mathbf{v})$, and expend the snake with a balloon force as $\mathbf{F}_b=E_{cst}(\mathbf{v})$, then $\mathbf{\Sigma F}=\mathbf{F}_b+\mathbf{F}_{im}$. After γ iteration steps, we can get new points ($\mathbf{u}_t$, $\mathbf{w}_t$) from ($\mathbf{u}_{t-1}$, $\mathbf{w}_{t-1}$). Fig.1 shows one of some results by the described model.

4. INTEREST POINT DETECTOR

4.1 Contour code

We select two *tracking windows* with different scales in input space $A'_m \subset I_i(\mathbf{x})$ and output space $A'_n \subset I_j(\mathbf{x})$ for the *kernel*'s geometric characteristics. The *tracking windows* is made around every selected *kernel* that is the center of mass in A'_m. If any more than two of elements of A'_m in A'_n are decided, the corresponding position of the *kernel* in $I_j(\mathbf{x})$ can be estimated. If some of corresponding points of the *kernels* are found in A'_n, we can also estimate all motion states in this new frame by prior knowledge. Corresponding elements in A'_m and A'_n are recognized by simple directive *chain codes* with their neighborhoods. Fig.2 gives a example of this principle: **left** shows there some points around a point P. If some of these points are understood, the corresponding point P in **right** can be located easily.

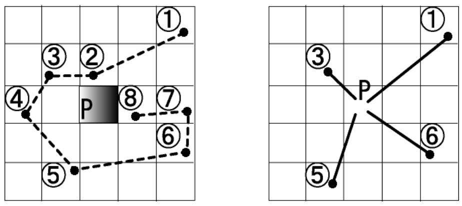

Fig. 2: Principle of chain code.

4.2 Behavior of orientation

We build up a transform relation $H(\mathbf{x})$ in a local window about a point $\mathbf{x}$ as follows:

$$H(\mathbf{x}) = \Re(\mathbf{x}) ** \left[(\bigtriangledown I(\mathbf{x})(\bigtriangledown I(\mathbf{x}))^T\right]. \tag{5}$$

where $I(\mathbf{x})$ is intensity function of the image, $**$ is the convolution operation. $\Re(\mathbf{x})$ is a weight mask to weight the derivatives over the window. This matrix captures the local structure. The eigenvectors of this matrix are the principal curvatures of the auto-correlation function. Its rank one indicates an edge and rank zero a homogeneous region. Two significant values indicate the presence of an interest point. We consider a cost function $M(\mathbf{x})$:

$$M(\mathbf{x}) = t_1 det[H(\mathbf{x})] + t_2 trace[H(\mathbf{x})]. \tag{6}$$

where t_1, t_2 are preset constants. The interest points are decided by selecting a threshold about $M(\mathbf{x})$. Fig.3 shows the detected results for an image by this method in our experiments.

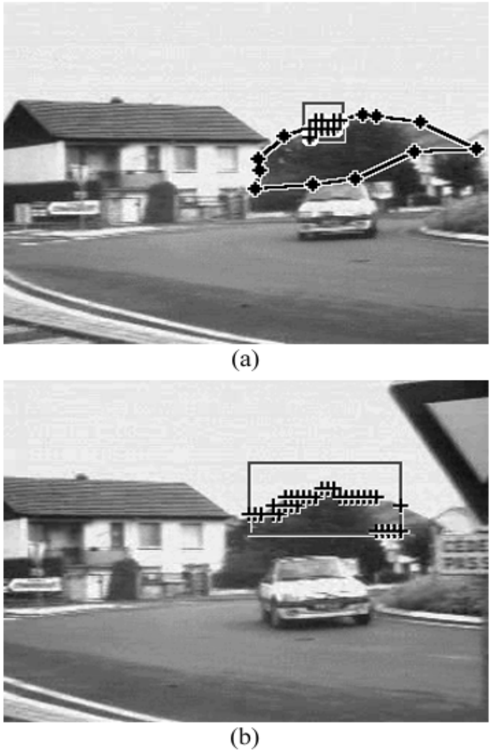

Fig. 3: Interest points of an selected image area. (a): the input frame; (b): the output frame.

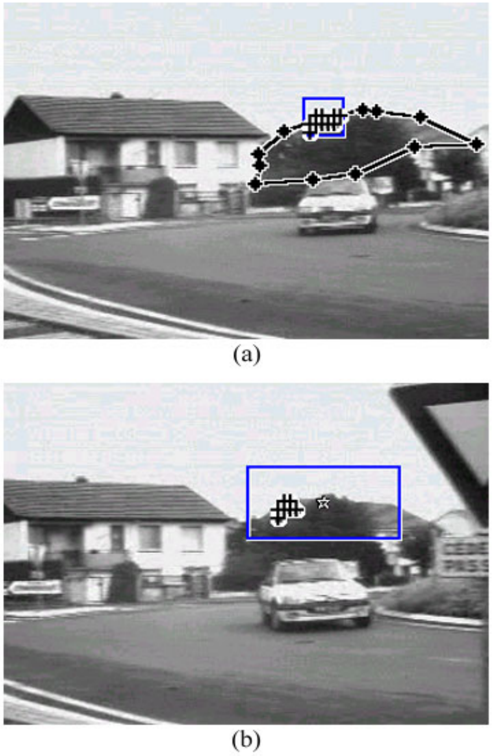

Fig. 4: Detecting correspondences: 10 points in (a) correspond with 7 points in (b). The star in (b) labels an kernel position from the snake model.

5. CORRESPONDENCE DETECTOR

After the process of interest point detection, we have to decide the relation of those interest points between A'_m and A'_n because the interest points cannot tell us anything about their correspondences. Let A_m=$\{\mathbf{x}_m\}$ be the set of all interest points in the first starting image that is input state space, and $A_n = \{\mathbf{y}_n\}$ the interest points in the second image that is output state space. Let $\mathbf{c}_{mn}$ be a vector connecting points $\mathbf{x}_m$ and $\mathbf{y}_n$. Two points $\mathbf{x}_m$ and $\mathbf{y}_n$ can be considered potentially corresponding if their distance satisfies the assumption of maximum velocity, $|\mathbf{x}_m - \mathbf{y}_n| \leq d_{max}$, d_{max} is the maximum distance a point may move in the time interval between two consecutive images. Two correspondences of points $\mathbf{x}_m\mathbf{y}_n$ and $\mathbf{x}_k\mathbf{y}_l$ are termed consistent of $|\mathbf{c}_{mn} - \mathbf{c}_{kl}| \leq c_{dif}$, c_{dif} is a preset constant derived from prior knowledge. Consistency of corresponding point pairs will increases the probability that a correspondence pair is correct. We Determine the sets of interest points $A_m \subset A'_m \subset I_i(\mathbf{x})$, $A_n \subset A'_n \subset I_j(\mathbf{x})$, and construct a data structure as follows:

$$[\mathbf{x}_m, \{\mathbf{c}_{m_i}, P_{m_i}\}, (NV^*, NP^*)], i = 1, 2, \ldots, n. \tag{7}$$

where P_{mn} is defined as the probability of correspondence of points $\mathbf{x}_m$ and $\mathbf{y}_n$, NV^*, and NP^* are special symbols indicating that no potential correspondence was found.

We initialize the probabilities P_{mn} as $P^{(0)}_{mn}$ as follows:

$$P^{(0)}_{mn} = \frac{1}{1 + k_p\omega_{mn}}(1 - P^{(0)}_{(NV*,NP*)}). \tag{8}$$

where $P^{(0)}_{(NV*,NP*)}$ is the initialized probability of no correspondence, k_p is a constant and

$$\omega_{mn} = \sum_{\Xi}[I_m(\mathbf{x}_m \pm \Xi) - I_n(\mathbf{y}_n \pm \Xi)]^2. \tag{9}$$

Ξ defines a neighborhood for image match testing –a neighborhood consists of all points $(\mathbf{x}+\Xi)$, Ξ is defined as a symmetric neighborhood around $\mathbf{x}$. We iteratively determine the probability of correspondence of a point $\mathbf{x}_m$ with all potential points $\mathbf{y}_n$ as a weighted sum of probabilities of correspondence of all consistent pairs $\mathbf{x}_k\mathbf{y}_l$, $\mathbf{x}_k$ are neighbors of $\mathbf{x}_m$ and the consistency of $\mathbf{x}_k\mathbf{y}_l$ is evaluated according to $\mathbf{x}_m$, $\mathbf{y}_n$. A quality q_{mn} of the correspondence pair is defined as

$$q^{(s-1)}_{mn} = \sum_k\sum_l P^{(s-1)}_{kl}. \tag{10}$$

where s denotes an iteration step, k refers to all points $\mathbf{x}_k$ that are neighbors of $\mathbf{x}_m$, and l refers to all points $\mathbf{y}_l \in A_n$ that form pairs $\mathbf{x}_k\mathbf{y}_l$ consistent with the pair $\mathbf{x}_m\mathbf{y}_n$.

The probabilities of correspondence are updated for each point pair $\mathbf{x}_m$, $\mathbf{y}_n$.

$$\hat{P}^{(s)}_{mn} = P^{(s-1)}_{mn}(k_a + k_b q^{(s-1)}_{mn}). \tag{11}$$

where k_a and k_b are preset constants. They deal with the convergent speed of P_{mn}. Normalize

$$P^{(s)}_{mn} = \frac{\hat{P}^{(s)}_{mn}}{\sum_j \hat{P}^{(s)}_{mj}}. \tag{12}$$

Those interest points that hold high probabilities that obviously differ from those interest points without correspondences. Repeat (10) (11) and (12) until the $P_{mn}^{(s)} > P_{thr}$ (threshold) is found for all points $\mathbf{x}_m$, $\mathbf{y}_n$.

Fig. 5: Selected frames: the 11th frame is up and the 15th down.

6. EXPERIMENTAL RESULTS

In this section, we give some of results based on the presented method. We use two frames of a video image to execute our experiments. The 11th and the 15th frames are selected for the input/output as shown in Fig.5. The active contour model has been shown in Fig.1. The results detecting interest point and correspondence are shown in Fig.3 and Fig.4 respectively. Fig.6 gives the amplified effects. There are different points in input frame and output frame, it means that more than one points in A_m corresponding to one points in A_n. We can use the method in Section 4.1 to recognize them easily. In this experiment, the parameters are $N = 12$, α=0.6, β=0.0, γ=25, $\mathbf{F}_{im}=g(\bigtriangledown I(\mathbf{x}))$, $\mathbf{F}_b=0$, $g(.)$ is a function. We select pixel distance in a time interval D_{max}=37.5 pixels, D_{dif}=1 pixel, s=10, P_{thr}=0.70. The correspondence points are decided with errors less than 5 pixels because our calculating image unit is 5 pixels. Fig.6 show us very good detecting results. The results have been good enough to analysis the motion features.

7. CONCLUSIONS

In the work, we proposed three processes of the method to tracking a static object from a moving platform when an camera is moving and an object is static: active contour model, which uses estimates of kernel points at the contour position, a interest point detection, which uses

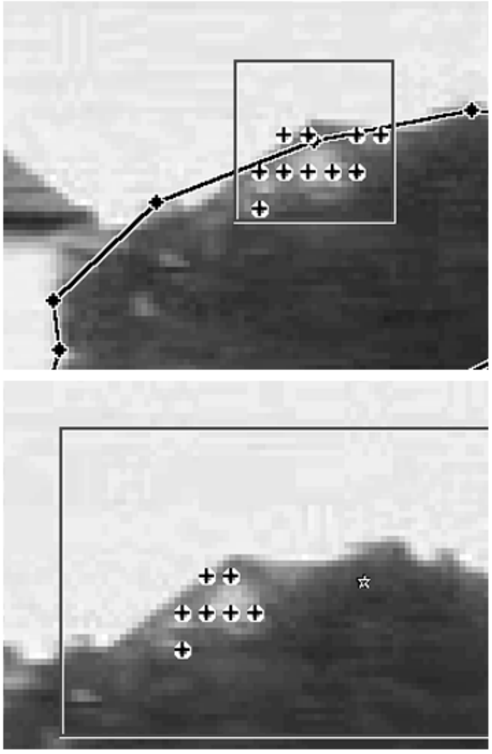

Fig. 6: The amplified images. The up and down images are the amplified (a) and (b) in Fig.4 respectively.

kernel′s area, along with measures of auto-correlation transform, as the system input measurement, and a probabilistic relaxation method, which uses correspondence pairs detection, as the former system output and the next system input. This approach takes us another advantage to occlusion problem, because the targets/objects can be estimated by parts of their features.

In future work, we hope to rationalize the selection visual cues used for object tracking based on prior image knowledge, and to give the system some of learning abilities.

REFERENCES

[1] A. Kumar, A. Tannenbaum, and G. Balas, “Optical flow: a curve evolution approach,” *IEEE Trans. Image Processing*, vol. 5, no. 5, pp. 598-610, 1996.

[2] R. Green, D. Gray, and J. Powers, *Artificial Vision*, Academic Press, London, 1997.

[3] M. Kass, A. Witkin, and D. Terzopouls, “Snakes: active contour models,” *Int'l J. Computer Vision*, vol. 1, no.4, pp. 321-331, 1988.

[4] G. Carneiro, and A. Jepson, “phase-based local features,” *Computer Vision - ECCV 2002, Part I*, pp. 283-296, 2002.

[5] W. Yu, K. Daniilidis, and G. Sommer, “Approximate orientation steer ability based on angular Gaussians,” *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 193-205, 2001.

[6] D. A. Forsyth, J. Ponce, *Computer Vision*, Prentice Hill, USA, 2003.

[7] W. Kropatsch, R. Klette, F. Solina, and R. Albrecht, *Theroetical foundations of computer vision*, Springer Wien, New York, 1996.