Maximum likelihood approach to the problem of simultaneous contouring and tracking

Yu. V. Slynko
Moscow Institute of Physics and Technology (State University), Russia, 141700 Dolgoprundy, Institutsky per. 9
E-mail: *UrasSl@yandex.ru*

Abstract – *Here we propose the method for tracking of the selected object in video sequence in the absence of a-prior information about background evolution, as well as size and a form of an object. Under such conditions a contouring of an object has to be performed prior to a tracking. Maximum likelihood method is employed for efficient contouring. Estimation of likelihood function, based on residual functions of all parts of frame, is proposed. We implemented an iterative algorithm for maximizing the estimated likelihood function. Developed algorithm showed high performance and stability in a wide spectrum of videos.*

Keywords - *Tracking, Contouring, Maximum Likelihood Method*

1. INTRODUCTION

Precise tracking of selected objects in a sequence of video frames is an important task for surveillance and security systems, intellectual systems of vehicle control and robotics. Besides, a good solution for objects tracking in videos can be used as a starting point in development of more complex systems for video analysis.

Proposed algorithm was designed for controllable cameras with ability for large. This fact imposes certain constraints on the functionality of the program. First of all, the algorithm should not be based on the information about the background in the frame. Fast movement of the camera, especially in case of large zooming, will not allow for accumulation of information about the background. Plus, background can be non-stationary (a sea surface, for example). Another constrain is the absence of information about the size and the form of an object. This constrain comes from the fact that the algorithm was designed for systems, where operator have to control many parameters in real-time, such as camera direction, for example. In this case operator is unable to set a precise contour of an object.

At the moment, it is still challenging to propose a good algorithm for solving a tracking problem task, since there is no universal solution to it proposed so far. Two approaches can be considered. First approach is based on the detection of the foreground, followed by classification of the objects in the frame (selection of a particular object amongst all objects of the foreground). Second approach is based on the localization of the object selected in the previous frame in the subsequent frames.

In the first case, algorithms usually work with stationary cameras and have many limitations. They are unable to track still objects and objects positioned against changing background. Such tasks are common for surveillance and monitoring systems. The algorithm in this case uses a well characterized stationary background. Most of the commercial products available on the market nowadays refer to this category. Review of such methods can be found in [1].

For tracking objects positioned against a more complex background, algorithms using filtering (accumulation) are usually applied. The most commonly used representatives of this class are the predictive algorithms ([2], [3]). Among more complex ones in this class are particle filtering ([4], [5]) and kernel-based algorithms ([6], [7]).

The main problem of filtering is the initiation of tracking, i.e. processing of the first few frames. There information have not being accumulated yet. This is especially relevant when algorithm has to control the camera. If it fails to track the object, it will disappear from the field of view in a few frames.

For solving the task of simultaneous tracking and contouring the algorithms of foreground classification and tracking must be strongly interconnected. One of the examples of such mutually-connected algorithms can be found in [8]. However, it has a strong limitation on the objects form, it has to be similar to a rectangle.

Here an algorithm is proposed, which simultaneously solves two tasks: it tracks the moving object and determines its shape in the absence of a priori information on the object size and shape and the evolution of the background.

2. THEORETICAL BACKGROUND

Maximum likelihood method was chosen to determine the shape and movement of the object simultaneously. In absence of a priori information about an object the method is optimal.

Frame f_t of time t consists of a finite number of parts, pixels (parts consisting of few pixels could be used for acceleration of the calculations; further on, they will be referred to as "parts"). We consider that each part completely belongs to an object or a background. An object choice is performed by the operator and consists of choosing a part i_0, which is known to be a part of this object.

Likelihood function can be written as following

$$J(I) = \prod_{i \in I} p(i \in object) \prod_{i \notin I} p(i \notin object), \qquad (1)$$

where i – number of the parts in the frame, I – set of parts that belong to an object, $p(i \in object)$ - probability of the fact that part i belongs to an object.

In order to calculate the object shift, the maximum of likelihood function has to be determined

$$I = \arg\max_I J(I). \qquad (2)$$

The main problem consists in determination of the probability $p(i \in object)$. This probability depends on the shift of an object and a background, as according to the task conditions the only difference between an object and a background is determined by the shift vector.

Residual function as a function of shift $(\delta x, \delta y)$ for each part and the entire object can be calculated as (details of calculations can be found in [9])

$$\tilde{C}_i(\delta x, \delta y) = \sum_{l,k \in P_i} \left(f_t(l,k) - f_{t-1}(l - \delta x, k - \delta y)\right)^2,$$
$$\tilde{F}(\delta x, \delta y) = \sum_{i \in I} \tilde{C}_i(\delta x, \delta y), \qquad (3)$$

where P_i - set of pixels which belong to part i.

We consider calculated residual functions as estimations of the true functions C_i and F. Minima of these functions are estimations of shifts of each part or the whole object. One could select parts of object for which the shift is close to a shift of i_0. But such method would be very unstable due to large errors in the estimation of a single part shift.

The calculation of the probability $p(i \in object)$ as the probability of coinciding minima of C_i and F will give more accurate result. Then

$$p(i \in object) = \sum_{\delta x, \delta y} p_i(\delta x, \delta y) p_F(\delta x, \delta y) \qquad (4)$$

or

$$p(i \in object) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_i(\delta x, \delta y) p_F(\delta x, \delta y) d\delta x d\delta y \qquad (5)$$

where $p_i(\delta x, \delta y)$ and $p_F(\delta x, \delta y)$ - probabilities of the fact, that shifts of the part and an object, respectively, are equal to $(\delta x, \delta y)$.

These probabilities can be estimated from residual functions. Distribution of the shift error estimations for each part is replaced with normal distribution

$$p_i(\delta x, \delta y) = \frac{1}{2\pi\sqrt{\det(\Gamma_i)}} e^{-\frac{1}{2}(\vec{x} - \vec{x}_i)^T \Gamma_i^{-1} (\vec{x} - \vec{x}_i)},$$

where

$$\vec{x} = (\delta x, \delta y), \qquad (6)$$

$$\Gamma_i = \begin{pmatrix} \sigma_{ix}^2 & \rho_i \sigma_{ix} \sigma_{iy} \\ \rho_i \sigma_{ix} \sigma_{iy} & \sigma_{iy}^2 \end{pmatrix}$$

and similarly for F.

Parameters of this distribution $\vec{x}_i, \sigma_{ix}, \sigma_{iy}$ and ρ_i can be calculated from residual function (see APPENDIX 1 for details). Integrating (5) leads to

$$p(i \in object) = \frac{\exp(\frac{B}{2A})}{2\pi\sqrt{-A}}, \text{ where}$$
$$A = 2\rho_i \rho_F \sigma_{Fx} \sigma_{Fy} \sigma_{ix} \sigma_{iy} + \sigma_{Fx}^2((\rho_F^2 - 1)\sigma_{Fy}^2 - \sigma_{iy}^2) + \sigma_{Fx}^2((\rho_i^2 - 1)\sigma_{iy}^2 - \sigma_{Fy}^2),$$
$$B = (\sigma_{Fy}^2 + \sigma_{iy}^2)(x_F - x_i)^2 - 2(\rho_F \sigma_{Fx} \sigma_{Fy} + \rho_i \sigma_{ix} \sigma_{iy})(x_F - x_i)(y_F - y_i) + (\sigma_{Fx}^2 + \sigma_{ix}^2)(y_F - y_i) \qquad (7)$$

Given shape of object I its shift can be calculated as

$$(\widetilde{\delta x}, \widetilde{\delta y}) = \arg\min_{(\delta x, \delta y)} \tilde{F}(\delta x, \delta y) \qquad (8)$$

3. IMPLEMENTATION DETAILS

In order to solve the problem of the object tracking with simultaneous contouring it is necessary to find maximum of likelihood function (1). The main obstacle here is the fact that residual function F depends of the contour I of an object. The problem could be solved accurately with an exhausting search method. However it is very time-consuming due to its long calculation time.

In order to facilitate the calculation and stability of the method we define parts as squares of 8 by 8 pixels. The following method is then proposed. Initial estimation of I is performed by any

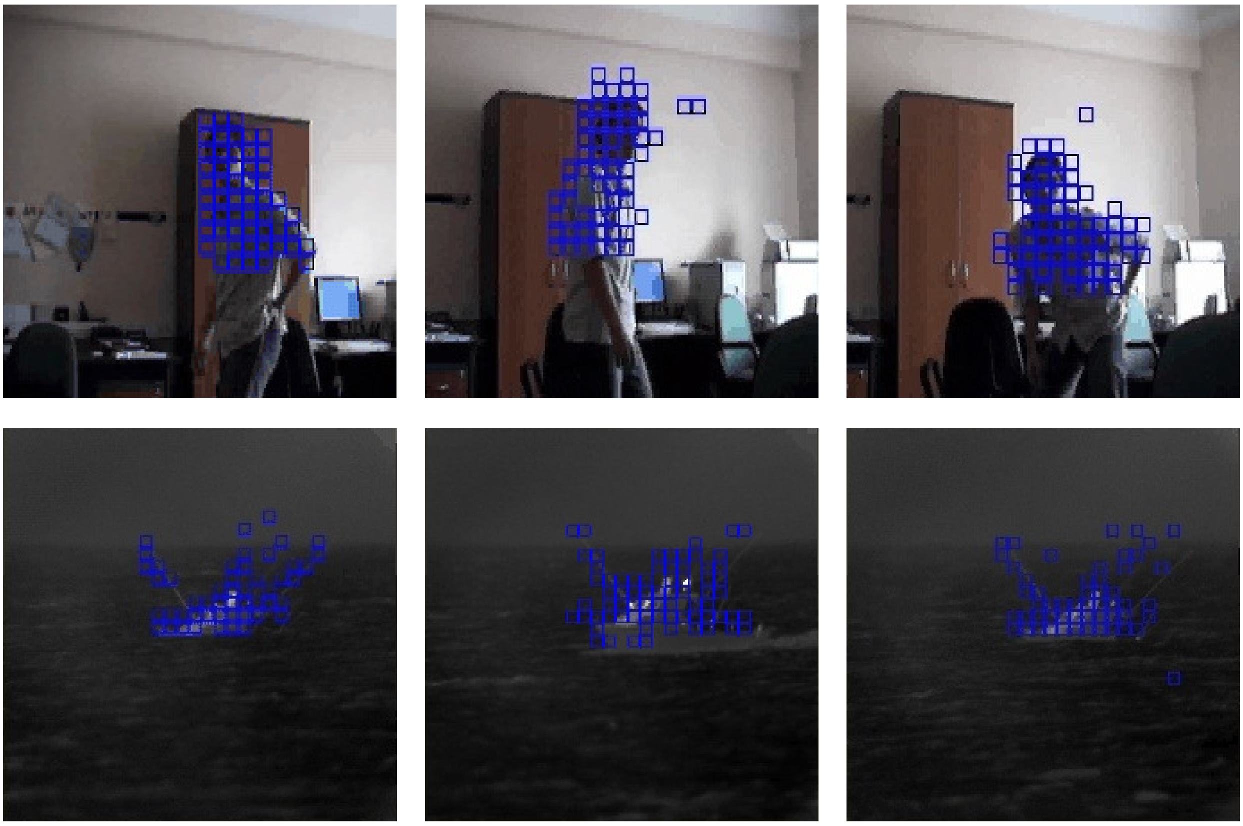

Fig. 1. Selected frames of the test videos.

conventional clusterization algorithm. This estimation is then refined iteratively based on the likelihood function.

The task of clusterization algorithm is to split frame into two clusters – the object and the background based on the shifts of all separate parts. It is implemented in a shift space with K-mean method.

Given initial I local maximum of likelihood function can be found iteratively.

As a first step residual function $\tilde{F}(\delta x, \delta y)$ for contour I is calculated. Based on this function the probabilities $p(i \in object)$ and $p(i \notin object)$ can be found for each part. If $p(i \in object) > p(i \notin object)$ the part belongs to the object, otherwise to the background. Thus a new contour of the object is created. The iteration process continues (i.e. for new contour function $\tilde{F}(\delta x, \delta y)$ calculated etc.) until contour I does not change any more. In practice such approach allows to achieve the local maximum in 3-4 iterations. This maximum is usually close or coincides with the global maximum.

4. RESULTS AND CONCLUSIONS

The algorithm for simultaneous tracking and contouring of moving objects was developed. The proposed algorithm was tested on a large number of distinctive video sequences. Special attention has been paid to the cases with a non-stationary background and a fast moving camera (i.e. fast background movement).

Two examples are shown in Fig. 1. Top row: tracking of a human with a moving camera. Bottom row: tracking of a boat on the sea surface with large waves. Small details were also detected by the algorithm, e.g. fishing rods and antennas on the tower of a boat. Parts, selected as objects, are shown in blue squares.

The estimation of the accuracy of the tracking algorithm was performed on simulated video sequences, where the non-stationary background was overlaid with a moving object. Investigated videos had shift estimation RMSD less than 0.1 pixel.

Here we developed and tested the robust algorithm for simultaneous tracking and contouring of the moving objects, which works in the absence of an a priori information about the status of the object and of the background.

REFERENCES

[1] Fatih Porikli. “Achieving Real-Time Object Detection and Tracking Under Extreme Conditions.” *Journal of Real-Time Image Processing*, Springer, 2006

[2] C. Wren, A. Azarbayejani, T. Darell, and A. Pentland, “Pfinder: Real-time tracking of the

human body," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 19, 1997, pp. 780–785.
[3] Y. Boykov and D. Huttenlocher, "Adaptive bayesian recognition in tracking rigid objects," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. II, Hilton Head, SC, 2000, pp. 697–704.
[4] M. Isard and I. Blake, "Condensation – conditional density propagation for visual tracking," *Intl. J. of Computer Vision*, Vol. 29, 1998, pp. 5–28.
[5] N. Bouaynaya, W. Qu, and D. Schonfeld, "An online motion-based particle filter for head tracking applications," *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, 2005.
[6] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, Hilton Head, SC, , 2000, pp. 142–149.
[7] F. Porikli and O. Tuzel, "Object tracking in low-frame-rate video," *Proc. of PIE/EI - Image and Video Communication and Processing*, San Jose, CA, 2005.
[8] S. Avidan, "Ensemble tracking," in Proc. IEEE *Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.
[9] Yu. V. Slyn'ko, V. N. Lagutkin, and A. P. Luk'yanov. "Development and Investigation of Real-Time Robust Algorithms for Estimating the Parameters of Geometric Transformations of Video-Sequence Frames." *Journal of Communications Technology and Electronics*, Vol. 52, No. 3, 2007, pp. 332–338.

APPENDIX 1

Estimation of the distribution of the residual function C_i minima was performed as following.

Based on the assumption of the pixels' white normal-distributed noise a confidence region of a residual function minimum can be found [9]

$$\tilde{D} = \left\{\delta x, \delta y : \tilde{C}_i(\delta x, \delta y) < m_n + B \cdot \sigma_n\right\}, \qquad (9)$$

where m_n - value of a minimum of the residual function, $\sigma_n = \frac{m_n \sqrt{2}}{\sqrt{N_n}}$ - dispersion of the residual function values, N_n – number of pixels in area in which function C_i was calculated (for a part with a size of 8 by 8 pixels $N_n = 64$), B - threshold, which determines a confidence probability of the error.

The distribution parameters of (6) can be calculated for this region as following

$$\vec{x}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \frac{1}{M(\tilde{D})} \sum_{\delta x, \delta y \in \tilde{D}} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix},$$

$$\begin{pmatrix} \sigma_{ix} \\ \sigma_{iy} \end{pmatrix} = \frac{1}{M(\tilde{D})} \sum_{\delta x, \delta y \in \tilde{D}} \begin{pmatrix} (\delta x - x_i)^2 \\ (\delta y - y_i)^2 \end{pmatrix}, \qquad (10)$$

$$\rho_i = \frac{1}{M(\tilde{D})} \frac{1}{\sigma_{ix}\sigma_{iy}} \sum_{\delta x, \delta y \in \tilde{D}} (\delta x - x_i)(\delta y - y_i),$$

Calculation of parameters for F can be performed similarly.