Real-time mean-shift based tracker for thermal vision systems

by G. Bieszczad* and T. Sosnowski**

*Military University of Technology, Electronic Department, Kaliskiego 2, 00-908 Warsaw, Poland
**Military University of Technology, Institute of Optoelectronics, Kaliskiego 2, 00-908 Warsaw, Poland

Extended Abstract /Poster presentation/

Real-time object tracking is the critical task in many computer vision applications such as surveillance, object-based video compression, or driver assistance. The most challenging issues encountered in visual object tracking are cluttered background, noise, occlusions and appearance change off tracked objects. The whole process of object tracking is made with component called tracker. In typical visual tracker we can distinguish two major components. One is responsible for target representation and localization, while second is responsible for filtering and data association with respect to object dynamics. Visual tracker have to cooperate with some other components. To start tracking procedure interesting object have to be detected. Detection algorithm have to point interesting object out for the tracking algorithm. This component is not always necessary. There are systems where interesting object is pointed by the system operator. Results of tracking are also used in some reasoning and decision taking component.

Tracking algorithms can be divided in four broad categories:

1. Gradient-based methods locate target objects in the subsequent frame by minimizing a cost function.
2. Feature-based approaches use features extracted from image attributes such as intensity, color, edges and contours for tracking target objects.
3. Knowledge-based tracking algorithms use apriori knowledge of target objects such as shape, object skeleton, skin color models and silhouette.
4. Learning-based approaches use pattern recognition algorithms to learn the target objects in order to search them in an image sequence.

Gradient-based methods like SSD (Sum of Squared Differences) evaluates target transition by finding changes between two consequent frames. Changes are estimated with gradients in space and time. This method has relatively low computational complexity but in practice is usable when small differences between two frames are assured. There is a need to add some special routines to make this algorithm immune to occlusions and lumination changes.

Knowledge-based methods demands excessive knowledge database about objects appearance and dynamics. They have high computational complexity and are used rather in off-line systems. In return this methods provide high accuracy.

Learning-based approaches unlike previous method class, do not use previously collected knowledge base because they learn objects properties on the spot. This method is also combined with detection algorithm because there is no provided prior information about which object should be tracked. This methods are roboust but the usage in real time systems is limited due to high computational power demands.

One of the most robust algorithm used in object tracking systems is color-based kernel density estimation, introduced in , which belongs to feature-based class. What is important, feature space is not limited to color data. It can employ any other feature space extracted from image with success. Localization is obtained by comparing and matching areas of image using estimated features. This procedure is usually made using Mean-Shift mode estimation. This approach gained more attention among research community due to its low computational complexity and robustness to appearance change, and evolved to many variations. Adaptation in this method is simpler than in learning-based methods, cause it employs only model update.

Thus exist many techniques for tracking objects but most of them was implemented in color vision systems. Tracking algorithms for thermal vision systems are not well investigated jet. This article will treat about adopting Mean-Shift tracking algorithm to thermal vision image sequences. It will emphasize on hardware implementation issues like computing demands for real-time operation and possibilities of task parallelization. Article will also put under consideration accuracy of thermal-based tracker, and used in this application mean-shift algorithm convergence.

Digital thermal image is build from elementary objects called pixels of certain values that represent temperature. This pixel set is oriented in perpendicular Cartesian coordinate system. Object from our area of interests is represented as a subset of pixels. To characterize the object, first a feature space have to be chosen. The reference target model is represented by its Probability Density Function (PDF) in the feature space. For example, the reference model can be chosen to be the temperature PDF of the object. The reference model is associated to the point **0** = (0,0) in the spatial domain - center of the reference object. The target candidate localized at point **y** is represented by its own temperature PDF. Both probability density function is estimated from m-bin histogram. Histogram is not the best PDF estimation but it's satisfying in this application, and it can be efficiently calculated in hardware like proposed in . The target and the model has certain size. What is more, in the practice, pixels placed farer form the center are affected by occlusions, background interference and other noises. To avoid this undesirable effect the kernel *k(r)* is added. In result probability of feature *u* of object is now calculated with formula below:

$$p_u(y) = c_p \sum_{i=1}^{n_p} \left[k\left(\left\| \frac{x_i - y}{h} \right\|^2 \right) \delta \left(b(x_i) - u \right) \right] \tag{1}$$

where:

$p_u(\mathbf{y})$	- probability of feature u in object centered at point y
$b(\mathbf{x}_i)$	- function associating pixel x_i to appropriate feature (here to histogram bin)
$k(r)$	- kernel profile
h	- bandwidth – constant that determinate object size (algorithms area of interest)
$\delta(x)$	- kronecker delta function
c_p	- normalization constant

To find new location of object in following frame, there is a need to find the most similar target object to the model. Similarity is obtained by the special coefficient called Bhattacharyya coefficient which is calculated from the eq. (2).

$$\rho(y) \equiv \rho[p(y), q] = \sum_{u=1}^{m} \sqrt{p_u(y) q_u} \tag{2}$$

where:

$p(y)$	- probability density function of features in object centered at point y
q_u	- probability density function of features in model object.

To accommodate comparisons among various targets, this distance should have a metric structure. To achieve this, distance between two probability density functions is calculated using Bhattacharyya coefficient from equation (3).

$$d(y) = \sqrt{1 - \rho[p(y), q]} \tag{3}$$

This statistical measure have some desirable properties like:

- It has metric structure
- Has geometric interpretation – it is a cosine of angles between two m-dimensional vectors p and q.
- It uses temperature as a feature space, therefore its invariant to scale and rotation.

Minimizing (3) is made by maximizing (2). Searching of new target starts from position of the target in the previous frame - y_0, and its neighborhood. To reduce algorithms computing complexity, the linearization of Bhattacharyya coefficient with Taylor expression around y_0 point was performed:

$$\rho[p(y), q] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{p_u(y_0) q_u} + \frac{1}{2}\sum_{u=1}^{m} p_u(y)\sqrt{\frac{q_u}{p_u(y_0)}} \tag{4}$$

Localization is obtained by searching for the coordinates of similarity function maximum. The search of maximum is done with mean-shift procedure. This iterative procedure moves from location $\mathbf{y_0}$ to the most similar location in area of interest placed at $\mathbf{y_1}$ Next the PDF is calculated for the new location $\mathbf{y_1}$. The localization is repeated until difference between estimation of two following points is smaller then given threshold. For real-time application, constrain in number of searching iteration is needed.

The presented method was simulated for thermal vision sequences using MATLAB software. The evaluation of thermal based prediction showed that it handles acceptable prediction accuracy but needs some algorithm trimming. Computational complexity is relatively low so algorithm can be used in highly time constrained systems. Some parts of tracking system like histogram calculating module or communication interfaces were implemented in hardware.

REFERENCES

[1] G. Hager, P. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, IEEE Trans. Pattern Anal. Mach. Intell. 20 (10) (1998) 1025-1039.

[2] M. Isard, A. Blake, Contour tracking by stochastic propagation of conditional density, in: European Conference on Computer Vision, 1996, pp. 343-356.

[3] G. Cheung, S. Baker, T. Kanade, Shape-from silhouette of articulated objects and its use for human body kinematics estimation and motion capture, in: IEEE Conference on Computer Vision and Pattern Recognition, vol.1, 2003, pp. 7784.

[4] O. Williams, A. Blake, R. Cipolla, Sparse Bayesian learning for efficient visual tracking, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1292-1304.

[5] D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non-rigid objects using mean-shift, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, 2000.

[6] G. Bieszczad, Implementation of picture histogram calculation module in FPGA, SECON2007 conference, Warsaw, Poland.

[7] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Machine Intel., vol. 24,no. 5, pp.603–619, 2002.