Mean-Shift Visual Tracking with NP-Windows Density Estimates

Chuan Zhao
zhao@robots.ox.ac.uk

Ian Reid
ian@robots.ox.ac.uk

Robotics Research Group,
Department of Engineering Science,
Oxford University

The mean-shift algorithm is a robust and easy method of finding local extrema in the density distribution of a data set. It has been used successfully for visual tracking in which the target is modelled using a colour histogram, and the image window with best matching histogram is sought. The estimation of these colour distributions in mean-shift visual tracking is essential since the goodness of convergence largely depends on them, and so if the accuracy of the distribution estimation can be improved, the tracking performance will also potentially improve, and we would reduce the risk of losing the target from accumulation of bias. However, a histogram, used in the standard mean shift algorithm, is potentially a poor estimate of the underlying colour distribution: it is not invariant to the image scale, the number of histogram bins or the number of samples, and this can have an adverse affect on the speed and accuracy of convergence of the mean-shift algorithm.

The key contribution of our paper is to apply a general non-parametric PDF estimation method, namely NP-Windows [3], to replace the histogram estimating the colour probabilities in mean-shift visual tracking. We show, in particular, that this leads to faster and more accurate convergence, and by making use of a (separable) Gaussian kernel rather than the more usual Epanechnikov kernel, that the actual run-time can be reduced as well.

NP-windows is a method for estimating the underlying probability density of a signal which is more accurate than a histogram for a particular class of data set, namely samples from a band-limited signal. It uses an interpolation scheme which fits piecewise functions to the signal samples, calculating the PDF by accumulating the contribution of each piecewise function. The main process consists of three main steps: first, between each pair of neighbouring samples (x_i, y_i) and (x_i', y_i'), the coefficients of a polynomial interpolation $(y = Y(x))$ are calculated; second, the PDF of y, $p_y(\cdot)$ is calculated analytically by transforming a uniform distribution on x through the transformation $Y(x)$; finally, bins at an arbitrary precision are populated from the $p_y(\cdot)$ for each pair of samples in order to generate the PDF for the whole target region. Although the last step generates the PDF by populating bins, the accuracy of this depends only on the accuracy of the piecewise representation, not the number of samples nor the number of bins. In our visual tracking case, the signals we are dealing with are images, and the piece-wise approximations come from bilinear interpolation across 4-neighbourhoods.

Comaniciu's original exposition of using mean-shift for visual tracking [2] made use of an Epanechnikov kernel to construct the mean-shift vector because of its optimality to minimise the average global error between the estimate and the true density, and its computational efficiency since the kernel evaluations involve only pairs of neighbouring points within distance smaller than the bandwidth rather than all pairs of points. Moreover, it leads to a particularly simple and beautiful form for the mean shift vector. Nevertheless, subsequent work [1] has demonstrated that the use of a Gaussian kernel can have better performance than the Epanechnikov kernel. In particular, use of the Gaussian kernel to construct the mean-shift vector can decrease the number of iterations required for convergence. Furthermore, since the Gaussian kernel is separable, there are computational gains in using it in preference to the Epanechnikov kernel; *viz*, the 2D kernel density estimate can be computed efficiently using two independent 1D Gaussian symmetrical masks, one in the x direction and one in the y direction.

The computational gains of this approach are not immediately obvious when considering convergence of the standard mean shift, because there, each iteration is almost trivial, so performing a few less iterations doesn't have a measurable effect on the total run time of the optimisation. However, in our case, each iteration requires calculation of the histogram using NP-Windows, which introduces considerable additional computational effort per iteration. We show, therefore, significant gains in using the separable Gaussian in preference to the Epanechnikov kernel, as well as improved accuracy.
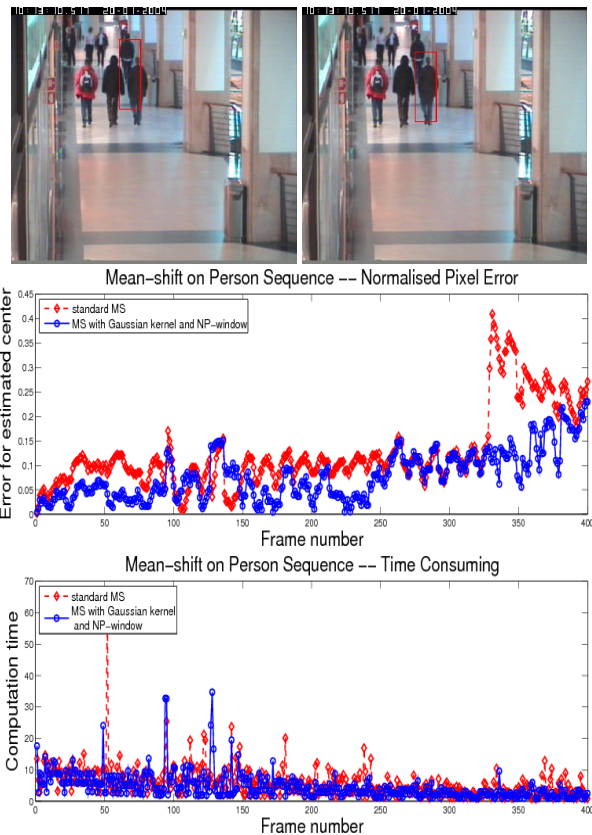

Figure 1: Tracking results of shopping center sequence at frame 350: standard MS (top-left) MS with Gaussian kernel and NP-windows (top-right); (middle) normalized pixel error: red, dashed line is MS and blue, solid line is MS with Gaussian kernel and NP-windows; (bottom) Run-time

Our improved mean-shift algorithm, therefore, simply replaces the Epanechnikov kernel with a separable Gaussian, and performs histogram estimation with NP-Windows. It is summarised in full in the paper.

An example tracking result is shown in the figures below, comparing the performance of standard mean-shift tracking, with our modified algorithm. Similar performance is observed on other sequences, in which we observe, on average, faster run-times, fewer iterations and smaller errors when using our modifications compared with standard mean-shift. We conclude that the use of NP-windows helps to improve the accuracy and slightly reduces the iteration times in Mean-Shift visual tracking.

As previously noted Kadir, NP-Windows is a general technique for histogram estimation from signal samples, and we therefore believe that its use in *any* visual tracking algorithm using colour distributions would be beneficial.

[1] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. PAMI*, 24:603–619, 2002.

[2] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE CVPR*, volume 2, pages 142–149, 2000.

[3] Timor Kadir and Michael Brady. Non-parametric estimation of probability distributions from sampled signals. Technical Report 2283/05, Robotics Research Laboratory, University of Oxford, 2005.