Object Tracking using Mean-shift Embedded Active Contours

Kyung-tai Kim\textsuperscript{1}, Jae Sik Chang\textsuperscript{2}, Hangjoon Kim\textsuperscript{2}, Han ku Lee\textsuperscript{1}, Eun Yi Kim\textsuperscript{1}

\textsuperscript{1}Dept. of Internet and Multimedia Eng., Konkuk Univ., Seoul, South Korea
\textsuperscript{2}Dept. of Computer Engineering, Kyungpook National Univ., Daegu, South Korea

Abstract. Active contour models have already been successfully used for object detection and tracking based on their ability to effectively describe curve and elastic properties. However, despite the successful use of active contour models, various technical problems still remain. The major problem is that their performance is very sensitive to the location of the initial curve. To alleviate this problem, a novel active contour model called mean-shift embedded active contours (MEAC) is proposed and applied to real-time object tracking. The main idea of the MEAC is to move the initial curve near the object before curve deformation. In the proposed method, the tracking is achieved by evolving the contour from frame to frame by minimizing some energy function evaluated in the narrow band around the curve. Then the MEAC performs a curve evolution by two steps: curve localization step and curve deformation step. When given the initial curve of the current frame from the object contour of the previous frame, the initial curve is firstly localized near the target object using mean-shift algorithm, and then it is deformed using a level set method. The MEAC is evaluated with several synthetic image sequences and natural sequences, then the results show that it improves the performance in terms of the convergence speed and the accuracy of the tracking results.

Keywords: Object tracking, Active contours, Mean-shift algorithm, textured or colored object

1 Background

Due to wide-ranging applications from video surveillance to the robot control, there has been substantial research in the area of object tracking. During the last decade, active contour models (ACMs) have been increasingly interested by many researchers. ACMs are effective to describe the elastic properties of the non-rigid objects, so that can provide the detailed analysis of the shape deformation during the objects are moving through the whole video sequences.

However, despite of their successful use in the object tracking, many research results indicate that the reliance of the performance on the initial curve is still prevalent problems in ACM-based tracking methods: the search efficiency for the target object is determined by the location of the initial curve and they have limited ability for the maximum range of distance between initial curve and object contour. The farther initial curve from object, induced by larger object motion, enforces the
wider search region to find the object contour. Unfortunately the curve deformation is a
time consuming task and the wider region is potential to include more local optima.
Accordingly, when the initial curve is far from the object, the convergence involves a
much heavier computational cost, plus errors are induced, such as noise and holes that
have similar features to an object contour. Moreover the methods fail to track the
highly active objects with large movement.

2 Proposed Method

In this paper, mean-shift embedded active contours (MEAC) are developed for real
time object contour tracking robust to the location of the initial curve.

In the proposed MEAC, the curve evolution is achieved by two steps: the curve
localization step and curve deformation step. In the first step, a mean shift algorithm
is used to move the initial curve to minimize the posterior energy without changing
the shape and scale of the curve. A level set method is then used to deform the curve
toward the object contour that minimizes the posterior energy.

Fig. 1 shows the scheme of the proposed MEAC. The blue solid lines are object
contour, while the red dotted lines are the evolved curve. The initial curve, shown in
Fig. 1(a), is localized near the object, as shown in Fig. 1 (b). The re-localized curve is
then deformed until it fines the object of interest.

![Fig. 1. Scheme of the proposed method. (a) Initial curve, (b) Localized curve, (c) Curve deformation.](image)

Let \( \tilde{\gamma}(p):[0,1] \rightarrow \mathbb{R}^2 \) be a closed planar curve. Then, \( g \) can be separated by \( \tilde{\gamma} \)
into the region enclosed by \( \tilde{\gamma}_e \), \( R_\tilde{\gamma} \), and its complement region, \( R^c_\tilde{\gamma} \). Thus, \( R_\tilde{\gamma} \) and
\( R^c_\tilde{\gamma} \) have a common boundary \( \tilde{\gamma} \), i.e. \( \tilde{\gamma} = \partial R_\tilde{\gamma} = \partial R^c_\tilde{\gamma} \), where \( \partial R \) is the boundary of
region \( R \).

**Curve Localization:** The objective of this step is to move the initial curve near the
object of interest without changing the shape and scale of the curve. Here, the
following energy function is used:

\[
E_m(\tilde{\gamma}) = \sum_{R_\tilde{\gamma}} (g_s - \mu)^T \Sigma^{-1} (g_s - \mu)
\]

To minimize \( E_m(\tilde{\gamma}) \), a mean shift algorithm based on adaptive search window is used,
which finds the object by seeking the mode of the object score distribution within the search window. In the present work, a Gaussian is used as the object score distribution at site $s$, which represents the probability of belonging to an object.

The algorithm iteratively shifts the center of the search window to the weighted mean until the difference between the means of successive iterations is less than a threshold.

**Curve Deformation:** After the initial curve is localized, the curve is deformed to find the optimal curve $\vec{y}$ that minimizes the posterior energy function using a level set method. As the curve deformation can be treated as a two-class discriminant analysis of the pixels on curve $\vec{y}$ and the memberships of the pixels are only dependent on their likelihood and prior probabilities, the computation area of the posterior energy can be reduced to the pixels on curve $\vec{y}$ as follows:

$$E_v(\vec{y}) = \sum_{\vec{y} \in \mathbb{R}^2} \left[ (g_s - \mu_i)^T \Sigma_i^{-1} (g_s - \mu_i) + \sum_{s \in \mathbb{C}} S_i (\omega_s (\vec{y}) = \lambda) \right] + \sum_{\vec{y} \in \mathbb{R}^2} \left[ -V + \sum_{s \in \mathbb{C}} S_s (\omega_s (\vec{y}) = \lambda) \right].$$  \hspace{1cm} \text{(2)}

To minimize the posterior energy function $E_v$, the steepest descent is taken with respect to $\vec{y}$.

Once the object contour is detected in the first frame, the initial curve in the current frame is then successively obtained using the object contour of the previous frame, so as to guarantee a temporal correspondence between the detected objects in successive frames. And then the initial curve is evolved until it finds the object contour. Then unlike previous active contours, the MEAC performs curve evolution by two steps: curve localization step and curve deformation step. The main idea of the proposed model is to move the initial curve near the object without shape and scale changes before the curve deformation. For this, we use a mean shift algorithm with adaptive search window. Then the proposed model reduces the computation cost and induces more accurate results, because the curve localization is simpler than curve deformation and jump over the local optima such as noise and holes.

3 Experimental results

To test the proposed model, we used several image sequences. The test data are classified into the textured objects sequences and colored object sequences, based on the used features to describe the target objects.

Fig. 2(a) shows the tracking of textured objects, where the background was composed of horizontal strips and the object was composed of diagonal strips. Then, the objects are described by a simple texture vector based on the directions of the (nonzero) intensity gradients in the neighborhood of a pixel. Fig. 2(b) shows the tracking result of colored object. For the photometric variable to describe the hand and the face, skin-color information was used as represented by a 2D-Gaussian model over chromatic color space.
As shown in Fig. 2, the proposed model was successful in tracking the hand and face throughout the entire sequences, regardless of the frame rates and the speed of the non-rigid objects.

To quantitatively evaluate the proposed model, we compared with other method and model in terms of speed and quality. Then, two methods are adopted for comparison: the general curve evolution method and Freedman’s method [1,2].

The general curve evolution method is implemented using level set method, where the same energy is used with the proposed MEAC. And the Freedman’s model, which is a recent remarkable active contour model for object tracking, finds the region where the sample distribution of the interior of the region most closely matches and it of the exterior of the region most mismatches the model distribution of the object, unlikely the proposed model that finds the region where include more pixels that have high density over the model distribution. These two methods are first compared with the proposed method, to prove the effectiveness of the MEAC which use two-step evolution scheme. Thereafter, to verify the soundness of the proposed energy function, the performance of the Freedman’s method is compared with one of the general curve evolution method that use the same energy with the proposed method.

Then, the Chamfer distance [3] was used as evaluation measure, which is defined as follows:

\[
C(F,G) = \frac{1}{3} \sqrt{\frac{1}{n} \sum_{i=1}^{n} v_i^2} \tag{3}
\]

where \(F\) and \(G\) are sets of pixels on the object boundaries detected by the proposed method and manually, respectively. In Eq. (3), \(v_i\) are the distance values from each point on \(F\) to the closest point on \(G\) and \(n\) is the number of points on the curve.

The quantitative comparison for these results is shown in Fig. 3. Fig. 3(a) shows the Chamfer distances for the three methods, and Fig. 3(b) shows the convergence speed in the three methods. In the three methods, the respective average times to process a frame in the image sequence are 0.469, 1.188 and 1.235, and the respective average
distances are 2.012, 8.372 and 10.538. With the proposed model, the Chamfer distance decreased more dramatically, allowing the stopping criterion to be reached after the fewer iterations than the general curve evolution and Freedman’s model. The curve re-localization of the proposed model, which moves only location of the curve without change of scale and shape of it, provides the search efficiency to track the contour of the non-rigid moving objects. Due to this, the proposed model induces the superior result quantitatively. And the general curve evolution was the faster converged to the less Chamfer distance value than the Freedman’s model, as the Freedman’s model needs computing time of the sample distribution of inside the curve to discriminate the object and background pixels while the proposed energy uses a manually selected parameter.

**Fig. 3.** The performance comparisons of the methods, (a) Chamfer distances, (b) Convergence times

As shown in this experiment, the superiority of our method was proven in the comparison between the proposed MEAC and the other two methods. And the comparison between the general curve evolution and Freedman’s method shows the effectiveness of the energy of the MEAC.
Acknowledgments. This research was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)"(NIPA-2012-H0301-12-3006)

Reference