# Active Segmentation and Adaptive Tracking Using Level Sets

Zezhi Chen[1] and Andrew M Wallace[2]
[1]School of Mathematics and Computer Sciences
[2]School of Engineering and Physical Sciences
Joint Research Institute in Signal and Image Processing
Heriot-Watt University, Edinburgh, UK, EH14 4AS
{Zc19, A.M.Wallace}@hw.ac.uk

**Abstract**

We describe algorithms for active segmentation (AS) of the first frame, and subsequent, adaptive object tracking through succeeding frames, in a video sequence. Object boundaries that include different known colours are segmented against complex backgrounds; it is not necessary for the object to be homogeneous. As the object moves, we develop a tracking algorithm that adaptively changes the colour space model (*CSM*) according to measures of similarity between object and background. We employ a kernel weighted by the normalized Chamfer distance transform, that changes shape according to a level set definition, to correspond to changes in the perceived 2D contour as the object rotates or deforms. This improves target representation and localisation. Experiments conducted on various synthetic and real colour images illustrate the segmentation and tracking capability and versatility of the algorithm in comparison with results using previously published methods.

## 1 Introduction

In this paper, we address the problem of segmentation and tracking of human subjects through video sequences, in which the subject is defined by an enclosing contour and a colour distribution within that contour, and the background may be static (fixed camera) or moving (panning camera) and defined by another colour distribution. In general, the colours within the foreground and background may change due to a different viewpoint or change of illumination. The work is founded on earlier work on mean-shift [5], level-set [2][7] and combined [3]methods to segment images and track deformable shapes in video sequences. In summary, there are three improvements over previous work.

The first process is segmentation on the first frame of the sequence to define the shape to be tracked. This uses an active segmentation (AS) algorithm based on level set methods and a multi-phase colour model. However, we have defined a general variational formulation which combines the Minkowski distance $L_2$ and $L_3$ of each channel and their homogenous regions in the index, as a change to the previous CVV model [1]. This method finds whole object boundaries that include different known colours, even in very

complex background situations, and shows improvement in synthetic data, in which the additive noise is non-Gaussian and asymmetric, and on real image data.

Second, we have developed an adaptive object tracking algorithm that combines AS and a mean shift tracking. The tracking algorithm has two phases. Assuming a current shape in a frame of index, *i*, then the mean shift algorithm can be used to find the most likely position of that same shape in frame $i+1$. Then, the AS algorithm deforms the contour to find a contour that better fits the data in the same $(i+1)^{th}$ frame. The approach is adaptive, in that it allows both deformation of the contour, and a change of the colour space model (*CSM*), the latter building on the work by Collins et al. [4] throughout the processing of a video. However, we sort the different *CSM*s using the Bhattacharyya coefficient which is an approximate measurement of the amount of overlap between the two distributions of foreground and background, instead of using the variance ratio measure of the distribution of likelihood values.

The third modification, when we obtain the boundary of a tracked object, is to use a kernel weighted by the normalized chamfer distance transform to improve the accuracy of target representation and localization. This replaces the more usual Epanechnikov kernel[6]. Comparative experiments show that our approach is more successful in tracking the object through video sequences, as both foreground and background colour distributions are better matched to the separated regions within the data.

# 2 Segmentation by Level Sets

## 2.1 Description of the Model

The basic idea in active contour segmentation is to evolve a curve, subject to constraints, in order to detect objects in the image. "*Let* $\Omega$ *be a bounded open subset of* $\mathbf{R^2}$, *with* $\partial\Omega$ *the boundary. Let* $\boldsymbol{I}$ *be a given image such that* $\boldsymbol{I} : \overline{\Omega} \longrightarrow \mathbf{R}$. *Let C(s) : [0, 1]* $\longrightarrow R^2$ *be a piecewise parameterized* $C^1$ *curve*" [1]. We make the following assumptions: 1) $\mathbf{I}$ is composed by a maximum of $\mathbf{M}$ regions $\Omega_i$; 2) the interface between the regions $\partial\Omega$ is regular. Our method also includes the minimization of an energy based function to perform segmentation. Describing image segmentation by a variational model increases the flexibility of the representation, allowing the future employment of additional features, such as shape knowledge, texture, motion vectors, etc. As implemented here, we assume a-priori knowledge of the colours of the object to be isolated. Given a $\mathbf{N}$-channel image $\mathbf{I}(\mathbf{I_1},\cdots,\mathbf{I_N})$, and a set of different colours/intensities $c=(c_1,c_2,\cdots,c_M)$. Then, $c_i,(i=1,\cdots,M)$ are vectors of length $\mathbf{N}$. The components of the foreground and background colours of the $k^{th}$ channel are $c^k_{fg}=(c^f_{k1},\cdots,c^f_{kR_f})$ and $c^k_{bg}=(c^b_{k1},\cdots,c^b_{kR_b})$ , $\mathbf{R_f}+\mathbf{R_b}=\mathbf{M}$. Figure 1 gives an illustration. We choose an energy formulation with the following form:

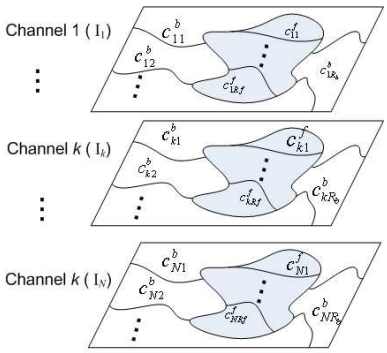


**Figure 1:** An image with N channels and a set of M different colours.

$$E(C)=\mu\cdot length(C)+\lambda_{\text{fg}}\cdot\iint_{\Omega_{\text{fg}}}F_{fg}(I(x,y),c_{\text{fg}})dxdy+\lambda_{\text{bg}}\cdot\iint_{\Omega_{\text{bg}}}F_{bg}(I(x,y),c_{\text{bg}})dxdy \quad (1)$$

where $C$ is the boundary curve of $\Omega_{fg}$ (shaded in Fig.1). $\Omega_{fg} = c^f_{k1} \cup \cdots \cup c^f_{kR_f}$ is the foreground (object) which is inside $C$, and the complement of $\Omega_{bg} = c^b_{k1} \cup \cdots \cup c^b_{kR_b}$ is the background which is outside $C$. Then, according to the bin-by-bin dissimilarity measurement - Minkowski distance [9], we use the mean of $L_2$ (the standard deviation) and $L_3$ (the third root of the skewness) in each channel to get the expressions:

$$F_{fg}(I(x,y),c_{fg}) = \sum_{r=2}^{3}(\prod_{q=1}^{R_f}(\frac{1}{N}\sum_{p=1}^{N}(w^f_q|I_p(x,y)-c^f_{pq}|^r))^{1/r})^{1/R} \tag{2}$$

$$F_{bg}(I(x,y),c_{bg}) = \sum_{r=2}^{3}(\prod_{q=1}^{R_b}(\frac{1}{N}\sum_{p=1}^{N}(w^b_q|I_p(x,y)-c^b_{pq}|^r))^{1/r})^{1/R} \tag{3}$$

where $c_i = average(I_p(x,y))$ inside the $i^{th}$ region. $\mu$, $\lambda_{fg}$ , $\lambda_{bg}$ and $w^{f,b}_i (i = 1,\cdots,N)$ are nonnegative weights for the regularizing term and the fitting term, respectively. This model is robust to symmetric and asymmetric noise (e.g. Gaussian and Gamma distributed noise). The optimal partition is obtained by minimizing the energy $E(C)$. *"The key idea is to evolve the boundary C to the boundary of the object from some initialization in direction of the negative energy gradient under the constraints from the image."*[7]

## 2.2 Level Set Formulation of the Model

For the level set formulation of the variational active contour model, we replace the unknown variable $C$ by the unknown variable $\phi$ , and follow [10], using the Heaviside function $H$, and the one-dimensional Dirac measure $\delta_0$ defined respectively by

$$H(z) = \begin{cases} 1 & , \quad if \quad z \geq 0 \\ 0 & , \quad if \quad z < 0 \end{cases} \qquad \delta_0 = \frac{d}{dz}H(z) \tag{4}$$

We express the terms in the energy $E$ in the following way:

$$E(C) = \int\int_{\Omega}(\mu \cdot \delta_0(\phi(x,y))\,|\nabla\phi(x,y)| + \lambda_{fg} \cdot F_{fg}H(\phi(x,y)) + \lambda_{bg} \cdot F_{bg}(1-H(\phi(x,y))))dxdy \tag{5}$$

In order to compute the associated Euler-Lagrange equation for the unknown function $\phi$ , our numerical simulations involve slightly regularized version of $H$ and $\delta_0$ , denoted here by $H_\varepsilon$ and $\delta_\varepsilon$ , as $\varepsilon \longrightarrow 0$. In this paper, we approximate the regularization of Heaviside by the complementary error function (erfc).

$$H_\varepsilon(z) = \frac{1}{2}erfc\left(-\frac{\sqrt{\pi}z}{\varepsilon}\right) \qquad \delta_\varepsilon(z) = H'_\varepsilon = \frac{e^{-\left(\frac{\sqrt{\pi}z}{\varepsilon}\right)^2}}{\varepsilon} \tag{6}$$

This is very similar to the procedure used by [1][2] and [10], but it has a bigger support interval, $(-\infty,+\infty)$. Minimizing $E(C)$ with respect to $\phi$ yields the following Euler-Lagrange equation for $\phi$ , parameterizing the descent direction by time, $t > 0$. The equation in $\phi(t,x,y)$ (with $\phi(0,x,y) = \phi_0(x,y)$ defining the initial contour) is:

$$\frac{\partial\phi}{\partial t} = \delta_\varepsilon(\phi)\left[\mu \cdot \nabla \bullet \left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \lambda_{fg}F_{fg} + \lambda_{bg}F_{bg}\right] \tag{7}$$

in $\Omega$ , and with the boundary condition $\frac{\delta_\varepsilon(\phi)}{|\nabla\phi|}\frac{\partial\phi}{\partial\vec{n}} = 0$ on $\Omega$ , where $\vec{n}$ denotes the normal at the boundary of $\Omega$ . Actually, $\frac{\nabla\phi}{|\nabla\phi|}$ is the unit (outward) normal, and the divergence of the normal $\nabla \bullet \left(\frac{\nabla\phi}{|\nabla\phi|}\right)$ is the mean curvature of the $\phi$.

## 2.3 Numerical Implementation

To solve this evolution problem, we use the level set method proposed by Osher [8]. We define an implicit function for $\phi$ using a signed distance. This function is positive on the exterior, negative on the interior, and zero on the boundary. Meanwhile, an extra condition of $|\phi| = 1$ should be satisfied. $\phi$ does not have to be a signed distance function; for example a Euclidean distance transform or Chamfer distance transform could be chosen as a level set function $\phi$. However, a signed distance function will increase the stability and quality of the evolution (especially if a vector field-based force and a force in normal direction are combined). This is because the signed distance is the path of steepest descent for the function. In order to improve numerical efficiency, we use a discrete form of the Hamilton-Jacobi (HJ) equation with high order ENO (Essentially Nonoscillatory) and WENO (Weighted ENO) accuracy and a Local Lax-Friedrichs (LLF) scheme. We also calculate the upwind derivative by using second order ENO scheme.

When working with level sets and Dirac delta functions, $\phi$ will no longer be a distance function (i.e. $|\phi| = 1$). $\phi$ can become irregular after some period of time. A standard procedure is to reinitialize the signed distance function to its zero-level curve. This prevents the level set function from becoming too flat, and can be seen as a rescaling and regularization. The reinitialization procedure is made by the following evolution equation:

$$\begin{cases} \psi_t = sign(\phi(t))(1 - |\nabla\psi|) \\ \psi(0, \bullet) = \phi(t, \bullet) \end{cases} \tag{8}$$

where $\phi(t, \bullet)$ is the solution $\phi$ at time $t$. Then the new $\phi(t, \bullet)$ will be $\psi$ , such that $\psi$ is obtained at the steady state of (8). The solution of (8) will have the same zero-level set as $\phi(t, \bullet)$ and away from this set, $|\nabla\psi|$ will converge to 1 [2].

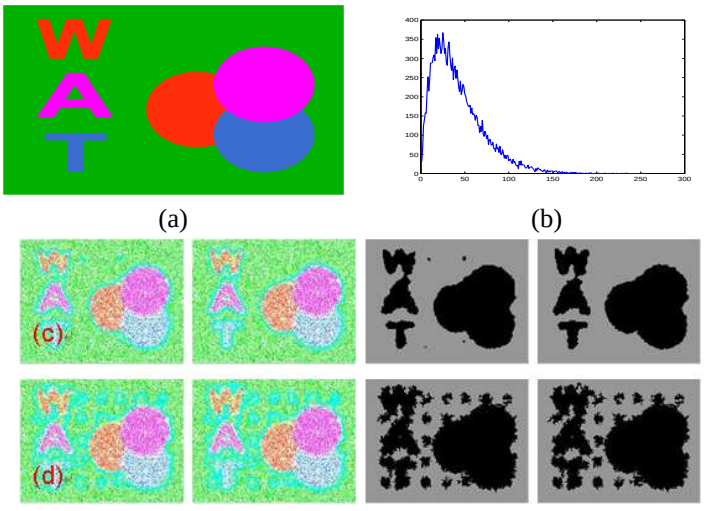


Figure 2: (a). Original synthetic image. (b). Gamma distributed noise. (c) and (d) show the results of iteration 40 and the final results by AS and CVV methods respectively; the noisy image is on the left, the associated piecewise-constant approximation on the right.

In all our numerical experiments, we generally choose the parameters: $\lambda_{fg} = \lambda_{bg} = 1$, $w_q^f = w_q^b = 1$. We use the approximations $H_\varepsilon$ and $\delta_\varepsilon$ of the Heaviside and Dirac delta functions ($\varepsilon = \Delta x = \Delta y$), in order to automatically detect interior contours, and to insure the computation of a global minimizer. Only the length parameter $\mu$ , which has a scaling role, is not the same in all experiments. If we have to detect all or as many objects as possible and of any size, then $\mu$ should be smaller. Otherwise, $\mu$ should be larger. To test the effect of the $L_3$ Minkowski distance in the energy function, we first add asymmetric noise (i.e. a Gamma distribution) to a synthetic image. Fig.2(a) shows the original synthetic image, (b) shows the noise distribution, (c) shows the results of the AS method, and (d) the CVV method. This shows the improvement of the AS over the CVV method if the noise is additive and asymmetric. In Fig.3, each method is applied to a real image with a coloured, striped texture. This shows that the AS can obtain the complete contour, but the CVV has breaks in the expected segmentation. AS only needs 87 iterations to converge to the optimal solution, but the CVV method takes 202 iterations.

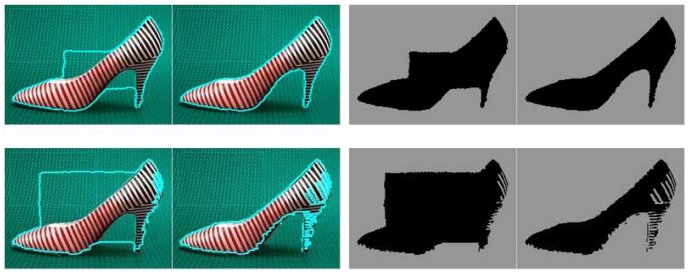

Figure 3: The comparison of the AS and CVV methods using a real image. The top figures are the results of iterations 60 and 87 by the AS method. The bottom figures are the results of iterations 60 and 202 by the CVV method. The original image is on the left and the associated piecewise-constant approximation is on the right.

We can also compare the accuracy of the AS method with that of the CVV method by calculating the energy of every evolution. Though energy formulation of the AS is different to that of CVV, and the initial value is different, we can compare the energy after normalization, because they should converge to the same global minimization, that is, $inf(E(C))$. For a perfect image and contours, $inf(F_{fg}) = inf(F_{bg}) = 0$, so $inf(E(C)) = \mu \cdot length(C)$. The comparison is shown in Figure 4. AS/CVV (Three colours) means we consider the three overlapping circles in Figure 2(a) as a single object and use the AS/CVV method. AS/CVV (one or two colours) has similar meaning. The experimental results show that the AS method only needs a small number of iterations to reach the minimum energy value. For example, for the object with three colours, the initial energy of the AS is bigger than that of CVV. After 25 iterations,

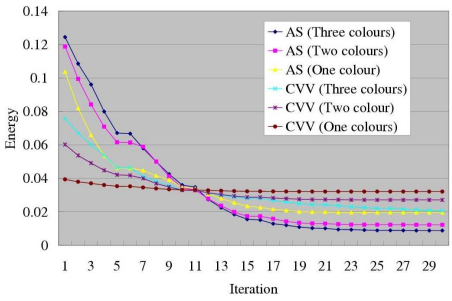


**Figure 4:** Comparisons of the energy evolution.

AS obtains a minimum, but the CVV method requires 150 iterations to obtain its final minimum.

# 3 Adaptive Object Tracking

## 3.1 Outline of the Adaptive Tracking Algorithm

The adaptive tracking algorithm is expressed as pseudocode,

- **Define the internal and external rectangles covering the object centroid at y0 in the first image.**
- **Sort *CSM*s by similarity distance criterion (Eq.10).**
- **Choose preferred *CSM*.**
- **Get active contour and $\phi$ of the tracked object by AS method.**
- **Repeat**
  - **Input the image $i$ (initial value $i = 1$).**
  - **Obtain the set of foreground and background pixels by $\phi$ .**
  - **Sort and choose preferred *CSM*.**
  - **Get the sets of constant colours by clustering using mean-shift segmentation.**
  - **Compute NCDT kernel using Chamfer distance transform.**
  - **Form model histogram, $q$, in the preferred colour space.**
  - **Fetch the next frame $i+1$.**
    - **Compute candidate histogram $p(y_0)$ in the preferred *CSM*s using NCDT-kernel**
    - **Find the optimum location $y_1$ of candidate using mean shift tracking algorithm.**
    - **Get the motion vector.**
  - **Translate the contours.**
  - **Update $\phi$ by AS method.**
  - $i = i+1$.
- **Until end of input sequence**

## 3.2 Selection of the Best Colour Space Model

In tracking an object through a colour image sequence, we shall assume that we can represent it by use of a discrete distribution of samples from a region in colour space, initially localised by a kernel whose centre defines the current position. Hence, we want to find the maximum in the distribution of a function, $\rho$, that measures the similarity between the weighted colour distributions as a function of position (shift) in the *candidate* image with respect to a previous model image. If we have two sets of parameters for the respective densities $p(x)$ and $q(x)$, the Bhattacharyya coefficient is an approximate measurement of the amount of overlap, defined by [6]:

$$\rho(y) = \rho[p(y), q] = \sum_{u=1}^{m} \sqrt{p_u(y) q_u} \tag{9}$$

The distance between two distributions can be defined as

$$d(y) = \sqrt{1 - \rho(y)} \tag{10}$$

Clearly the distance $d(y)$ lies between zero and unity, and obeys the triangle inequality. In a discrete space, $x_i, i = 1, 2, \cdots, n$ are the pixel locations of the model, centred at a spatial location $\mathbf{0}$, which is defined as the position of the window in the preceding frame

that we want to track. A function $b : R^2 \rightarrow 1,2,\cdots,n$ associates to the pixel at location $x_i$ the index $b(x_i)$ of the histogram bin corresponding to the value of that pixel. Hence a normalized histogram of the region of interest can be formed (using $q_u$ as an example)

$$q_u = \frac{1}{n}\sum_{i=1}^{n}\delta[b(x_i)-u], \quad u = 1,2,\cdots,m \tag{11}$$

where $\delta$ is the Kronecker delta function. Tracking success or failure depends primarily on how distinguishable an object is from its surroundings. If the object is very distinctive, it is easy to track. Otherwise it is hard to track. Normally, the features that best distinguish between foreground and background are the best features for tracking. The choice of feature space will need to be continuously re-evaluated over time to adapt to changing appearances of the tracked object and its background. To select the best colour space model (*CSM*), we sort the different *CSM*s using the Bhattacharyya coefficient which is an approximate measurement of the amount of overlap between the two distributions of foreground and background. For the first frame, we use a "centre-surround" approach to sample pixels from object and background. A rectangular set of pixels covering the object is chosen to represent the object pixels, while a larger surrounding ring of pixels of the rectangle is chosen to represent the background. For an internal rectangle of size $h \times w$ pixels, the outer margin of width $(\sqrt{2}-1)\sqrt{hw}/2$ pixels forms the background sample. The foreground and background have the same number of pixels if $h = w$. In all subsequent frames, it is the contour defined by the level set function, $\phi$, that defines the foreground for the adaptive model, so that no background is included. We use the distance criterion (10) to measure the similarity between the two histograms of the internal and external regions. The best colour space is selected by finding the *CSM* with maximum distance value. Each potential feature set typically has dozens of tunable parameters and therefore the full number of potential features that could be used for tracking is enormous. We construct 16 single frame *CSM*s from 5 different colour spaces (R.G.B, L.a.b, H.S.V, Y.I.Q/Y.Cb.Cr, C.M.Y.K). All the values of pixels are normalized to 0 to 255, yielding feature histograms with 16 or 256 bins.

Fig.5(a) shows a sample image with concentric boxes delineating the object and background. The similarity distances between foreground and background of each *CSM* are shown in Fig.5(b) and the set of all 16 candidate images after rank-ordering the feature according to the criterion (10) are shown in Fig. 5(c). The image with the most discriminative feature (best for tracking) is at the upper left. The image with the least discriminative feature (worst for tracking) is at the lower right.

## 3.3 Using a Kernel Based on the Normalized Chamfer Distance Transform

A radially symmetric kernel $K$ can be described by a 1D profile rather than a 2D (or higher order) image. The most popular choice for $K$ is the optimal Epanechnikov kernel that has a uniform derivative of $G = 1$ which is also computationally simple. However, in tracking an object through a video sequence and applying the mean shift algorithm to move the position of the target window, the bounds of the domain $R^2$ are altered on each successive application of the algorithm. There is no reason to suppose that the target has radial symmetry, and even if an elliptical kernel is used, i.e. there is variable bandwidth,

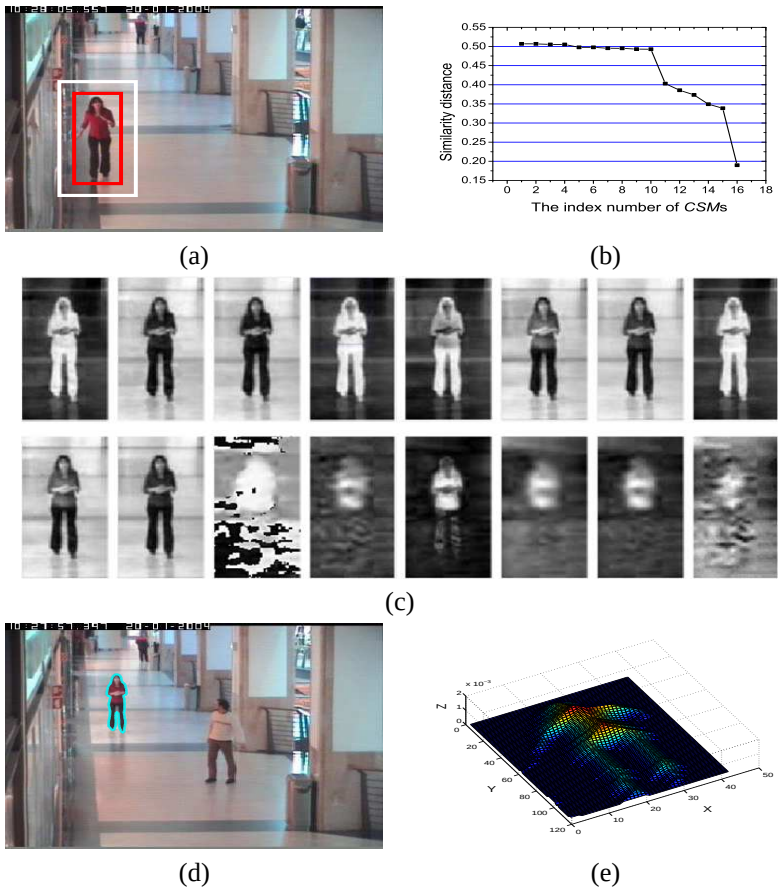


Figure 5: (a). A sample image with concentric boxes delineating the object and background. (b). The similarity distance of each *CSM*. (c). Rank-ordered 16 images. (d). AS segmentation result. (e). 3D NCDT kernel

the background area that is being sampled for the colour distribution will change. If the background is uniform this will not affect the colour *pdf*, and hence the gradient ascent will be exact. However, if it is not uniform, but varies markedly and in a worst case has similar properties to the target, as we shall see in the next section, then multiple modes will be formed in the *pdf* and the mean shift is no longer exact. Therefore, we use the normalised Chamfer distance transform (NCDT) rather than the true Euclidean distance, as it is an efficient approximation. The NCDT kernel better represents the colour distribution of the tracked target, yet retains the more reliable centre weighting of the radially symmetric kernels. This transform is applied to the target area, separated from the background by AS methods described in Section 2.3. Figures 5(d) and (e) show the AS segmentation and the NCDT kernel of Figure 5(a). We aim to show that this weighting increases the accuracy and robustness of representation of the *pdf*'s as the target moves, since it excludes peripheral pixels that occur within a radially symmetric window applied to successive frames. We are investigating the performance of the NCDT

transform to define the regions of interest and weight the colour densities in the video images. We assess whether the anticipated gain in excluding background pixels from the density estimates and weighting more substantially those more reliable pixels towards the centre of the tracked object will outweigh the possibility of forming false modes because of the shape of the NCDT. However, radially symmetric kernels may also produce false modes due to badly defined densities.

Figure 6 illustrates that the tracking algorithm can cope with dynamic deformation of the shapes and the changing positions of the targets in the various sequences, even when the camera pans so that both the foreground and background move in the camera coordinate system (Fig. 6(a)). All of these illustrations are from much longer sequences, included as supplementary material, typically more than a hundred frames. Fig.6(b) and (c) show that the tracking algorithm is very robust to clutter, and crossing objects. The car is occluded by a square object which has the same colour as the car in the third sample picture in (b), two people cross in the third sample picture in (c), yet the algorithm adapts the contour to track the non-occluded portion of the woman, then re-grows the contour as she re-emerges from behind the man. In each of the sequences, the rectangle in the first image defines the initial region, in which the object to be tracked is segmented. In Figure 6(c), we have also compared the use of the NCDT and Epanechnikov kernel, but in the latter case the tracker latches on the crossing individual.

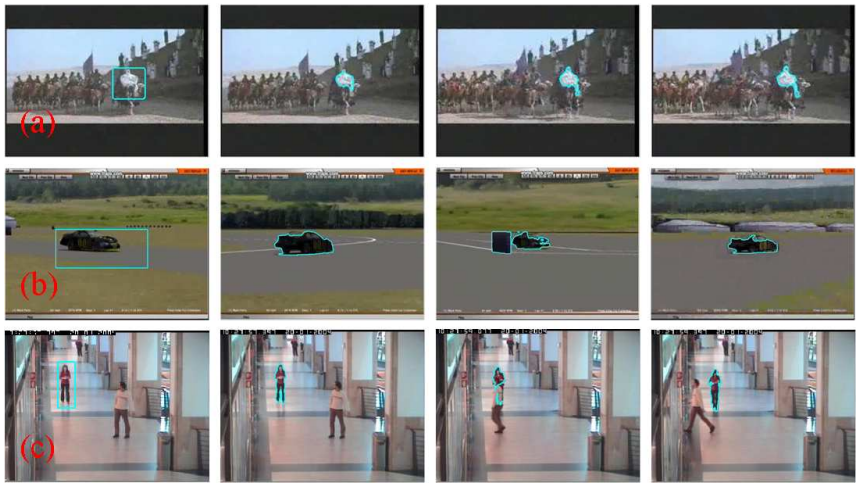


Figure 6: Tracking video objects and dynamics of deformation. The video sequences are supplied as supplementary material.

# 4 Conclusions

We have developed segmentation and tracking algorithms using a generalized active contour model. The object of interest can have a mixture of colours, but these are known before segmentation. For segmentation, the position of the initial curve can be anywhere in the image, and need not necessarily surround the object to be detected. However, if the

initial estimate is far from the true contour, it takes a long time to converge to the optimal solution. The segmentation is similar to the earlier CVV algorithm, but uses a slightly different cost function that deals better with noise that is asymmetric, and converges more quickly on sample image data. The adaptive object-tracking is a hybrid algorithm, combining level set methods with the mean shift tracking algorithm. Mean shift defines the translation in the next frame to accelerate the level set definition of the tracked contour. The algorithm is also improved by a Chamfer distance transform (NCDT kernel) and sorted CSMs to better detect and track objects. Several experiments have demonstrated the ability of the model to detect and track an object in movie sequences.

# References


[1] T. Chan, B.Y. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2):130–141, 2000.

[2] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2):266–277, 2001.

[3] J.S. Chang, E.Y. Kim, K. Jung, and H.j. Kim. Object tracking using mean shift and active contours. In *Proceedings of the 18th Int. Conf. on Innovations in Applied Artificial Intelligence*, pages 26–35, Bari, June 2005.

[4] R.T. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1631–1643, 2005.

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[6] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[7] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.

[8] R. Osher, S.and Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

[9] Y. Rubner, C. Tomasi, and L.J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[10] H.K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1):179–195, 1996.