Segmentation for Robust Tracking in the Presence of Severe Occlusion

Camillo Gentile
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

Octavia Camps and Mario Sznaier
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract

Tracking an object in a sequence of images can fail due to partial occlusion or clutter. Robustness can be increased by tracking a set of "parts", provided that a suitable set can be identified. In this paper we propose a novel segmentation, specifically designed to improve robustness against occlusion in the context of tracking. The main result shows that tracking the parts resulting from this segmentation outperforms both tracking parts obtained through traditional segmentations, and tracking the entire target. Additional results include a statistical analysis of the correlation between features of a part and tracking error, and identifying a cost function highly correlated with the tracking error.

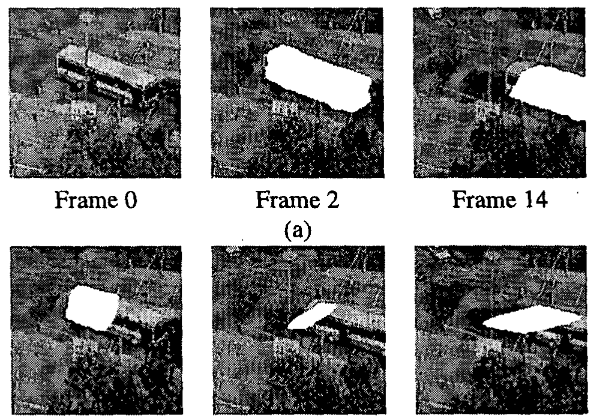
1 Introduction

Tracking a known object in a sequence of frames can fail due to occlusion or the presence of clutter. Robustness against these effects can be increased by using robust estimators [1, 8, 21, 26]. However, these estimators usually break down at above a 30% occlusion level[2]. This is illustrated in Figure 1(a)[1], where an affine transformation combined with a robust estimator was used to track a bus in a traffic sequence. As shown there, the algorithm begins to lose track of the target in Frame 14.

This effect can be traced to the fact that robust estimators treat occluding pixels as uniformly distributed outliers, neglecting the fact that occlusion tends to be clustered in small regions. Thus, intuitively one would expect that resiliency to occlusion could be improved by dividing the object into pieces which are tracked separately, along with the entire object, to find multiple transformations. The best global transformation is then selected by voting [30]. However, homogeneous pieces are more difficult to track than regions with distinctive properties such as texture or shape[2]. Thus, standard segmentations (see for instance [29, 15, 18, 17, 28]

[1]This sequence of traffic images was provided by Dr. Nagel at the Universitat Karlsruhe.

[2]This is closely related to the well known aperture problem.



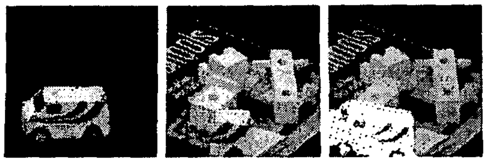
Frame 0 Frame 2 Frame 14
(a)

Frame 2 Frame 14 Frame 19
(b)

Figure 1. Tracking (a) using a robust affine transformation; (b) using homogeneous parts.

and references therein) do not necessarily result in parts leading to good tracking performance. This is illustrated in Figure 1(b) where the use of a set of homogeneous parts (an MDL-based segmentation [15] of the bus) leads to poor tracking starting in Frame 2. Motivated by this difficulty, in this paper we address the problem of how to divide the object into pieces to optimize tracking robustness to occlusion. Specifically, the contributions of the paper are:

- A statistical analysis of the correlation between features of a part and tracking error.
- Identifying a cost function that exhibits a higher degree of correlation with the tracking error than other indicators previously proposed.
- A segmentation algorithm specifically designed to make optimal use of the spatial information available to improve tracking robustness. This segmentation is obtained by combining this new cost function with the standard "snakes" framework [24, 29].



(a) (b) (c)

Figure 2. The PRA Algorithm.

2 Preliminaries

2.1 Notation and mathematical preliminaries

In the sequel we will denote by $\mathbf{x} = (x \quad y)^T$ the coordinates of a pixel; $I(\mathbf{x})$ the greyscale intensity at pixel $\mathbf{x}$; $I_x(\mathbf{x})$ and $I_y(\mathbf{x})$ the corresponding spatial derivatives; and by $\alpha \in R^6$ the parameters of an affine transformation $\mathcal{A}_\alpha$.

2.2 The tracking problem

For simplicity, in this paper we consider a prototype tracking algorithm based on estimating the transformation that maps the images of a given object between two consecutive frames in a sequence. However, our results can also be used in the context of tracking algorithms that exploit, in addition to spatial, temporal information [3, 19].

Assuming that there is little distortion, the mapping between consecutive frames can be considered to be affine [2, 4, 9, 14]. This fact can be exploited to efficiently solve the tracking problem by recasting it as [2]:

Problem 1 (Robust Affine Tracking:) *Given two image frames I^{f_1} and I^{f_2}, and a prototype object represented by a subset of pixels $P \subseteq I^{f_1}$, find an affine transformation $\mathcal{A}_\alpha : I^{f_1} \to I^{f_2}$ that minimizes:*

$$R(\mathcal{A}_\alpha) = \sum_{\mathbf{x} \in P} \rho\left\{I^{f_2}[\mathcal{A}_\alpha(\mathbf{x})] - I^{f_1}(\mathbf{x}), \theta\right\} \qquad (1)$$

where $\rho(., \theta)$ is a robust estimator [1, 2] that rejects outliers, controlled by the tuning parameter θ.

Problem 1 can be solved using gradient descent (in the sequel we will refer to this as the Benchmark Algorithm). However, as illustrated in Figure 1, this approach may fail in the presence of severe occlusion.

2.3 The Parts Reset Algorithm (PRA)

Since occlusion is a localized effect, robustness against severe occlusion can be improved by partitioning the prototype into a number N of parts and estimating N candidate affine transformations $\mathcal{A}_{\alpha_i}, i = 1, \ldots, N$, by tracking these parts [30]. A single transformation $\mathcal{A}_\alpha$ may be selected from the candidate transformations by a voting scheme. Further improvement can be obtained by evaluating the performance of each transformation at intermediate stages and resetting those that exhibit large errors, thus evading local minima. This effect is illustrated in Figure 2 (c) showing the tracking results for an algorithm based upon this idea. Here severe occlusion is simulated by cutting off a portion of the object *Van*, pasting it to a cluttered background and adding zero-mean additive white Gaussian noise with variance 5.

3 Good Features for Tracking

While the approach described in the previous section has the potential to handle substantial occlusion, it hinges upon determining a suitable set of parts to be tracked. Possible options span a very diverse spectrum from dividing the object image by using a simple grid, to segmenting the object into homogeneous regions, to dividing the object into its "functional" parts. While the first option is the simplest partition and the latter options are intuitively appealing, they are not necessarily the best partitions for the application being considered. In this section we analyze the correlation between different features of a part and tracking performance. Based on this analysis, we propose a new segmentation, designed to optimize tracking robustness.

3.1 Performance of several indicators

Several ways of assessing the "goodness" (in the sense of its ability to minimize the tracking error) of a part have been proposed in the literature, based on spatial derivatives, image Laplacian or the eigenvalues of the matrix

$$W = \sum_{\mathbf{x} \in P} \begin{bmatrix} x^2 I_x^2 & x^2 I_x I_y & xy I_x^2 & xy I_x I_y & x I_x^2 & x I_x I_y \\ x^2 I_x I_y & x^2 I_y^2 & xy I_x I_y & xy I_y^2 & x I_x I_y & x I_y^2 \\ xy I_x^2 & xy I_x I_y & y^2 I_x^2 & y^2 I_x I_y & y I_x^2 & y I_x I_y \\ xy I_x I_y & xy I_y^2 & y^2 I_x I_y & y^2 I_y^2 & y I_x I_y & y I_y^2 \\ x I_x^2 & x I_x I_y & y I_x^2 & y I_x I_y & I_x^2 & I_x I_y \\ x I_x I_y & x I_y^2 & y I_x I_y & y I_y^2 & I_x I_y & I_y^2 \end{bmatrix} \qquad (2)$$

Table 1 summarizes the most commonly used indicators [20, 7, 22, 23]. Here a higher value of the criterion indicates a part that is thought to be more suitable for tracking.

To establish the performance of these features as indicators of good tracking we conducted a set of experiments to find the correlation between the indicators values and tracking error. To this effect we considered a set of parts of varying size, shape, and texture cut from real images and ran a series of tests on each part. These parts bear interesting features used for comparison, namely large regions with homogeneous texture (such as the faces of a box) as well as regions with contrast texture (such as corners and holes). Each experiment consisted of the following steps:

Table 1. Indicators for good tracking commonly used.

Description	Definition
intensity variance	$var_I = \sum_{\mathbf{x}\in P} (I(\mathbf{x}) - \eta_I)^2$
gradient	$grad = \sum_{\mathbf{x}\in P} I_x^2(\mathbf{x}) + I_y^2(\mathbf{x})$
normalized gradient	$grad_n = \frac{1}{\zeta(P)} \sum_{\mathbf{x}\in P} I_x^2(\mathbf{x}) + I_y^2(\mathbf{x})$
normalized laplacian	$lap_n = \frac{-1}{\zeta(P)} \sum_{\mathbf{x}\in P} (I_{xx}(\mathbf{x}) + I_{yy})^2(\mathbf{x})$
max. abs. eigenvalue	$max_W = \max_i \lvert\lambda_i\rvert,\ Wv_i = \lambda_i v_i$
min. abs. eigenvalue	$min_W = \min_i \lvert\lambda_i\rvert,\ Wv_i = \lambda_i v_i$
norm eigenvalues	$norm_W = \sum_i \lambda_i^2,\ \ Wv_i = \lambda_i v_i$
ratio eigenvalues	$rat_W = -\frac{max_W}{min_W}$

Table 2. Test poses.

level of difficulty	translation (pixels)	rotation (radians)	number of poses
easy	3	0.00	8
moderate	7	0.00	8
difficult	3	0.15	16
challenging	4	0.18	16

1. Select a part P and a random background B with uniform grayscale distribution between 0 and 255.

2. Transform the prototype of the part P from the identity pose, $\mathcal{A}_0(\mathbf{x}) = \mathbf{x}$, to a test pose $\mathcal{A}_\alpha(\mathbf{x}) = \mathbf{x}'$ and "paste" it onto background B.

3. Corrupt the resulting scene with zero-mean additive white Gaussian noise with variance 5.

4. Run the tracking algorithm to compute an estimated pose, $\mathcal{A}_{\hat{\alpha}}$.

5. Find the corresponding ground–truth *tracking error* defined as:

$$d(\mathcal{A}_\alpha, \mathcal{A}_{\hat{\alpha}} | P, B) = \frac{1}{\zeta(P)} \sum_{\mathbf{x}\in P} ||\mathcal{A}_{\hat{\alpha}}(\mathbf{x}) - \mathcal{A}_\alpha(\mathbf{x})||_2 \quad (3)$$

where $\zeta(P)$ denotes the number of pixels of part P.

A total of 19200 experiments were performed using 40 parts, P_i pasted onto 10 128 $\times$ 128 random backgrounds B_j and under 48 different affine transformations α_k. These transformations correspond to the translations and rotations indicated in Table 2, where a translation τ summarizes the eight directions adjacent to a pixel, x0=-τ, y0=-τ; x0=-τ, y0=0; x0=-τ, y0=τ; etc; and the rotation denotes both clockwise and counter–clockwise motion. The range of displacements shown in Table 2 was chosen to cover problems ranging from easy to challenging.

The overall performance of a part P is obtained by summing over the 48 poses and ten backgrounds to compute the total error, $D(P)$ associated with it:

Table 3. Correlation between indicators and tracking error

indicator	Correlation
var_I	-0.3169
$grad$	-0.2766
$grad_n$	-0.0271
lap_n	-0.1705
max_W	-0.3228
min_W	-0.3143
$norm_W$	-0.3042
rat_W	-0.1915
proposed	**−0.7963**

$$D(P) = \sum_{j=1}^{10} \sum_{k=1}^{48} d(\mathcal{A}_{\alpha_k}, \mathcal{A}_{\hat{\alpha}_k} | P, B_j) \quad (4)$$

A potential problem when using equation (4) to assess the quality of part P is that a few outliers can significantly bias the cumulative performance of the part. To avoid this situation we proceeded as follows. Through the use of Kolmogorov–Smirnov test [16] we determined that, with probability ≥ 0.75, the distribution of the experiments yielding lowest values of the error is an F distribution (the ratio of two random variables with χ^2 distribution) with parameters $v_1 = 16$ and $v_2 = 4$. Since for this distribution, $F(d) = 0.95$ for the error value $d = 5.8$, all points above this value were considered outliers and assigned an error value of $d = 5.8$[3]. With this saturation the total error of a part ranges from 0 (perfect) to 2784 (poor matching).

The correlation coefficient $\sigma_{\mathbf{DJ}}$ between the 40-dimensional vectors $\mathbf{D}$ of tracking errors and $\mathbf{J}$ of indicator values defined in Table 1 are given in the top portion of Table 3. Unfortunately, all of them have small absolute value, indicating that the performance of these indicators as predictors of good tracking properties is rather poor.

3.2 Performance oriented indicator

To find an indicator more correlated with tracking performance we begin by examining the gradient with respect to the affine transformation parameters of equation (1) used to perform the search for the affine parameters:

$$\nabla_\alpha R = \sum_{\mathbf{x}\in P} \frac{\partial\rho(r(\mathbf{x}),\theta)}{\partial r(\mathbf{x})} \nabla_\alpha r(\mathbf{x}) = \quad (5)$$

$$\sum_{i=1}^{N} \sum_{\mathbf{x}\in P_i} \frac{\partial\rho(r(\mathbf{x}),\theta)}{\partial r(\mathbf{x})} \left[I_x^f(\mathbf{x}')\ I_y^f(\mathbf{x}')\ x I_x^f(\mathbf{x}')\ y I_y^f(\mathbf{x}')\ y I_x^f(\mathbf{x}')\ x I_y^f(\mathbf{x}') \right]^T$$

[3]Values about this threshold correspond to cases where there is a large mismatch between the actual and calculated poses. In this situation the numerical value of the error is more a function of the background than of the disparity between poses.

where $r(\mathbf{x}) = I^f[\mathcal{A}_\alpha(\mathbf{x})] - I^p(\mathbf{x})$, $P = \bigcup_{i=1}^{N} P_i$ and $P_i \cap P_j = \emptyset$ for $i \neq j$.

Equation (5) shows that, as expected, parts that have large spatial derivatives I_x and I_y as well as large momenta xI_x, yI_y, yI_x, and xI_y result in larger gradient of the objective function in (1) and thus in a faster convergence towards the optimum set of affine parameters. However, consistent numerical experience indicates that performance does not improve once the gradient components exceed a certain threshold. Based on these considerations we propose to use as an indicator of good tracking properties the *energy* of a part P, defined as:

$$\begin{aligned} e(P) &= \text{sat}\left[e_x(P), e_{sat}\right] + \text{sat}\left[e_y(P), e_{sat}\right] \\ &+ \text{sat}\left[e_{xx}(P), e'_{sat}\right] + \text{sat}\left[e_{yy}(P), e'_{sat}\right] \\ &+ \text{sat}\left[e_{yx}(P), e'_{sat}\right] + \text{sat}\left[e_{xy}(P), e'_{sat}\right] \end{aligned} \tag{6}$$

where

$$\begin{aligned} e_u(P) &= \sum_{\mathbf{x}\in P} \text{sat}\left[I_u^2(\mathbf{x}), I_{sat}\right] \\ e_{uv}(P) &= \sum_{\mathbf{x}\in P} |u - \eta_{uv}| \, \text{sat}\left[I_v^2(\mathbf{x}), I_{sat}\right], \\ \eta_{uv} &= \frac{1}{e_v(P)} \sum_{\mathbf{x}\in P} u \, \text{sat}\left[I_v^2(\mathbf{x}), I_{sat}\right] \end{aligned} \tag{7}$$

sat[.,.] is the saturation operator defined by

$$\text{sat}[c, c_{sat}] = \begin{cases} \frac{c}{c_{sat}} & \text{if } c \leq c_{sat} \\ 1 & \text{if } c > c_{sat} \end{cases}$$

and where η_{xx}, η_{yy}, η_{yx} and η_{xy} are used to center the momenta to render the energy coordinate independent.

The parameters I_{sat}, e_{sat}, and e'_{sat} are additional degrees of freedom that can be used to optimize the correlation between the tracking error E and the energy $e(P)$. For the set of 19200 experiments described at the beginning of the section, the optimal values of these parameters were found to be $I_{sat} = 3000$, $e_{sat} = 50$, and $e'_{sat} = 700$ respectively. The corresponding indicator $e(P)$ is highly correlated with the tracking error, with correlation coefficient $\sigma_{De} = -0.7963$[4]. Note that this value is substantially larger than the other entries in Table 3.

4 Object Segmentation for Tracking

In the last section we identified an indicator that exhibits a high degree of correlation with the expected performance of the part in a gradient–based tracking algorithm. Thus, tracking performance can be optimized by finding a partition of the object that minimizes this indicator value. In this section we describe how to accomplish this by incorporating the energy (6) of a part into a deformable model or snake framework [24].

4.1 Snake Description

A snake is an ordered set of points $S = [s_1, s_2, \ldots, s_n]$ that can form either open or closed contours. A snake segmentation algorithm moves the snake on the image grid seeking to minimize an energy function:

$$E = \sum_{i=1}^{n} E_{int}(s_i) + E_{ext}(s_i)$$

where the internal force E_{int} imposes continuity and smoothness constraints to avoid oscillations of the contours, and the external force E_{ext} attracts the snake to salient image features.

Let s be a point on the snake, U_s be the subset of points on the snake adjacent to s, and V_s be the set of parts defined by the snake that have s as a contour point. Then, the internal energy at the point s, $E_{int}(s)$, is defined as:

$$\begin{aligned} E_{int}(s) = &\alpha \max_{t\in U_s} \left\{(x_s - x_t)^2 + (y_s - y_t)^2\right\} + \\ &\beta \min_{t,u\in U_s} \left\{(x_t - 2x_s + x_u)^2 + (y_t - 2y_s + y_u)^2\right\} \end{aligned}$$

where the first term ensures that points on the snake do not get too far from each other, the second term penalizes high curvature contours, and α and β control the relative influence of the corresponding terms.

As discussed in the previous section, for a tracking application, the external force at the point s, $E_{ext}(s)$, should attract the snake towards enclosing parts with high trackability indicator values. Thus, the external force is defined as:

$$E_{ext}(s) = -\gamma \sum_{P\in V_s} e(P)$$

where $e(P)$ is the trackability indicator for part P as defined in equation (6) [5].

4.2 Snake Initialization

A snake in the form of a square grid placed on the object performs the initial segmentation, dividing it into a number of parts. (Other grids could serve as well, triangular, etc.)

[4] The negative value here indicates that larger values of the energy lead to smaller values of the tracking error. Computing the median rather than the average over all experiments for a part P in (4) also yields a high coefficient $\sigma_{De} = -0.7505$.

[5] For a correct implementation of the segmentation method, the contribution of each energy term must be normalized by dividing the term by the largest value in the neighborhood where the snake point can move: $\frac{E(s)}{\max_{t\in\mathcal{N}(s)} E(t)}$, where $\mathcal{N}(s)$ is the set of pixels in a neighborhood of s.

We attempt to find the global minimum of the energy through a greedy search. The search resembles the segmentation algorithm described in [29] through region competition based on snakes which guarantee closed parts, employing statistics inside the region rather than just information along the region boundary, and global optimization techniques based on an energy function.

4.3 Minimization of the Energy

The experimental results shown in section 3.2 indicate that a "good" segmentation for tracking can be obtained by maximizing the "energy" of the parts. However, simply maximizing this energy may lead to a segmentation composed of just a few large parts, or, in extreme cases, to a trivial solution with just one part. Clearly these solutions are undesirable in terms of robustness to occlusion. Moreover, as was the case in section 3.2, consistent experience shows that once the energy components of a part rise above a given threshold, little improvement in tracking performance is obtained by increasing them even further. Rather, performance can be improved by attempting to increase the energy components of the remaining parts above that threshold. Finally, note that the problem is non–convex and thus the minimization algorithm may get trapped in a local minima. To take these effects into account, rather than attempting to maximize the raw energy, we wil optimize the following filtered versions of the energy components:

$$E_{ext}(s) = -\gamma \sum_{P \in V_s} \hat{e}(P) \tag{8}$$

where

$$\begin{aligned} \hat{e}(P) = {} & f(e_x(P), e_{sat}, \frac{e_{sat}}{e'_{sat}}\lambda) + f(e_y(P), e_{sat}, \frac{e_{sat}}{e'_{sat}}\lambda) \\ & + f(e_{xx}(P), e'_{sat}, \lambda) + f(e_{yy}(P), e'_{sat}, \lambda) \\ & + f(e_{yx}(P), e'_{sat}, \lambda) + f(e_{xy}(P), e'_{sat}, \lambda) \end{aligned} \tag{9}$$

and

$$f(c, c_{sat}, \lambda) = \frac{1}{1 + \exp^{-\lambda(c - c_{sat})}}. \tag{10}$$

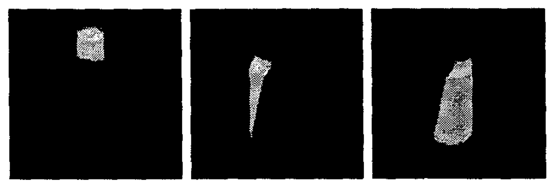
The parameter λ controls the shape of the filter and in the limit when $\lambda \to \infty$ it forces all the parts to have an energy above c_{sat} (since $f(c, c_{sat}, \infty) = 0$ for $c < c_{sat}$). By increasing λ from 0 to ∞ as the minimization proceeds we achieve an effect similar to simulated annealing [5], that minimizes the probability of converging to a local minimum. This process is illustrated in Table 4, showing three stages of the segmentation algorithm. Here we used the values $\alpha = \beta = \gamma = 1.0$, giving equal weight to the eight energy components.

The final segmentation is shown in Figure 3. Since the low energy content is concentrated in the upper portion of the object, the segmentation distributes this portion amongst the three parts. Although Part 2 and Part 3 share the lower portion, Part 3 grows much larger than Part 2. However as far as energy, they have about the same amount since the upper portion is divided about evenly between them.

The high energy of the lower portion forces the two parts which share it to be elongated. An alternative partition may have created an additional bottom part taking from Part 2 and Part 3 (for a total of four), with virtually no energy, thus creating a bad part. As we will show in section 5, an abundance of bad parts proves as detrimental to the PRA as a lack of good parts.

Table 4. Segmentation of *Cylinder* for three values of λ

λ=0.01 λ=1.0 λ=10.0



Part 1 Part 2 Part 3

Figure 3. *Cylinder* segmentation.

5 Tracking Results

In this section we report the results of a series of experiments comparing tracking performance of the PRA when using the proposed parts versus homogeneous parts obtained using a Mininum Description Length based algorithm [15]. For comparison purposes, we also include the tracking performance of the Benchmark algorithm.

These experiments were performed using six toy objects that exhibit interesting features, namely large regions with homogeneous texture as well as regions with contrast texture, supplemented with ten objects taken from real tracking sequences. As in section 3, for each of the 16 test objects we performed 480 tests:

1. 16 "challenging" poses of Table 2,
2. five cluttered backgrounds (similar to Figure 2 (b), containing objects with similar texture to the test objects, and the scene corrupted by zero–mean additive white Gaussian noise with variance 5), and

3. six translations of 50% occlusion (measured in the number of prototype pixels).

The resulting scores, ranging from zero to 2784, are shown in Table 5.

Table 5. Comparison between the tracking algorithms.

Object	(a) Benchmark	(b) PRA with homogeneous parts	(c) PRA with prop. parts
Block	2648.756	2112.312	1758.192
Log	2299.223	2077.602	1665.434
Box	2231.747	2029.325	1301.574
Cylinder	2012.668	2061.614	1186.690
Van	2537.833	2555.088	1404.820
Truck	1929.924	1580.622	788.190
Car	1908.622	1200.242	859.233
Car_2	2099.368	1902.682	994.273
Compact	2737.885	2446.042	2298.574
Compact_2	2603.391	2342.972	1908.958
Bus	2616.033	2119.049	1196.824
Race	2446.529	1802.315	1700.237
Wagon	2469.425	1881.417	1051.092
Wagon_2	2111.319	1833.022	900.994
Car_3	2084.411	2149.430	844.964
Cadillac	1965.081	1776.500	677.095

The PRA with the proposed parts (c) outperforms both the Benchmark algorithm (a) and the PRA with homogeneous parts (b) for each of the test objects.

- (c) outperforms (a) by up to 51% on an absolute scale (*Bus*, *Wagon*) and 190% on a relative scale (*Cadillac*), and
- (c) outperforms (a) on average by 36% on an absolute scale and by 95% on a relative scale.
- (c) outperforms (b) by up to 47% on an absolute scale (*Car_3*) and 162% on a relative scale (*Cadillac*), and
- (c) outperforms (b) on average by 25% on an absolute scale and by 69% on a relative scale.

It is worth noticing that (a) outperforms (b) in some occasions, illustrating again that using an homogenous segmentation can lead to worse results than not using parts at all.

Frame 2 Frame 14 Frame 19

Figure 4. Tracking the bus using the proposed segmentation.

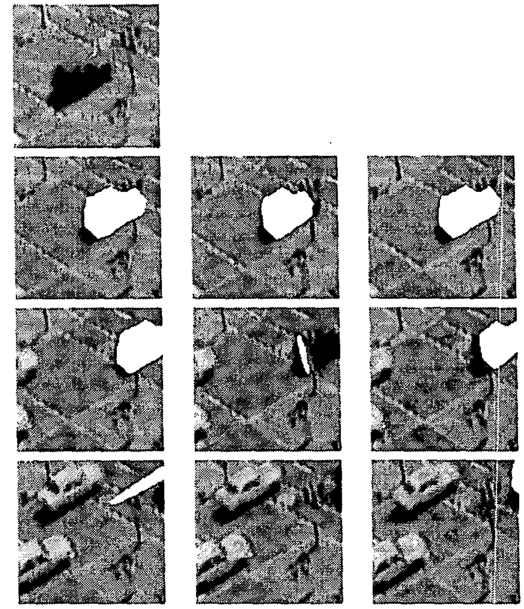
Figure 4 shows the results of using the proposed segmentation on the bus sequence. In this case the algorithm is able to successfully track the target throughout the entire sequence until the film ends at about 60% occlusion.

Frame 0

Frame 8

Frame 17

Frame 28



(a) (b) (c)

Figure 5. *Compact* sequence: (a) Object. (b) Homogeneous parts. (c) Proposed segmentation.

Finally, Figure 5 shows experimental results obtained with the *Compact* sequence[6]. In this case, using homogeneous parts leads to poor tracking after Frame 8: the mild occlusion caused by the traffic sign compresses a "bad" part into its unoccluded portion in order to minimize the error norm, and the object follows. The part continues to compress the object in the subsequent frames and eventually loses track of it completely. The traffic sign does not affect (a) and (c) at all, however by Frame 17 at about 30% occlusion (a) has begun to lose track of the object, while (c) successfully tracks the object throughout the entire sequence until it is last visible in Frame 28 at about 90% occlusion.

6 Conclusions

Many tracking algorithms used widely in the computer vision community deal with occlusion through a robust estimator. Such estimators fare well with moderate occlusion, but break down at above 30% occlusion level. To expand this range, in [30] we have proposed to track, in addition to the object, a set of parts. Intuitively, this idea exploits the fact that occlusion tends to be localized, and thus successful tracking can be accomplished as long as a few of these parts

[6] Additional experiments, omitted for space reasons, can be obtained contacting the authors.

exhibit less than 30% occlusion. However, as illustrated with several examples, successful application of this idea requires a suitable object segmentation. In this paper we have identified desirable properties (from a robust tracking standpoint) for the parts and proposed an energy function to obtain these parts by solving an optimization problem. Experimental results with both synthetic and real images show that, when used in a context of a tracking algorithm, these parts outperform those obtained using traditional segmentation methods.

References

[1] M.J. Black and P. Anandan. "A Framework for the Robust Estimation of Optical Flow." *IEEE ICCV*, May 1993: pp. 231-236.

[2] M.J. Black and A.D. Jepson. "Eigentracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation." *European Conf. on Comp. Vision*, April 1996: 1064, p. 329.

[3] A. Blake and M. Isard, "Condensation - condensation density propagation for visual tracking," *IJCV*,29, 1, pp. 5–28,1998.

[4] G.D. Borshukov, G. Bozdagi, Y. Altunbasak, and A.M. Tekalp. "Motion Segmentation by Multistage Affine Classification." *IEEE Trans. on Image Proc.*, Nov. 1997: 6(11) : pp. 1591-1594.

[5] N. K. Bose and P. Liang. *Neural Network Fundamentals.*, pp. 318-323. McGraw-Hill, Inc., 1993.

[6] F. Dornaika and C. Garcia. "Object Pose by Affine Iterations." *Int. Conf. on Image Analysis and Proc.*, Sep. 1997: 1310, pp. 478-485.

[7] J. Ferruz and A. Ollero. "Real-time feature matching in image sequences for non-structured environments. Applications to vehicle guidance." *J. of Intelligent and Robotic Systems: Theory and Applications*, : 28(1), pp. 85-123.

[8] E. Grossmann and J. Santos-Victor. "Performance evaluation of optical flow Estimators: Assessment of a new affine flow method." *Robotics and Auton. Sys.*, 1997: 21, pp. 69-82.

[9] N. Gupta and L. Kanal. "Gradient Based Image Motion Estimation Without Computing Gradients." *Int. Journal of Comp. Vision*, Feb.-Mar. 1997: 22(1): pp. 81-101.

[10] G. Hager and P. Belhumeur. "Efficient region tracking with parametric models of geometry and illumination." *IEEE Trans. on PAMI*, 1998: 20(10), pp. 1025–1039.

[11] C.Y. Huang, O.I. Camps, and T. Kanungo. "Object Recognition Using Appearance-Based Parts and Relations." *IEEE Conf. CVPR*, Jun. 1997: pp. 877-883.

[12] M. Irani, B. Rousso, and S. Peleg. "Computing Occluding and Transparent Motions." *Int. J. of Computer Vision*, February 1994: 12(1): pp. 5-16.

[13] S.X. Ju, M.J. Black, and Y. Yacoob. " Cardboard People: A Parameterized Model of Articulated Image Motion." *IEEE Conf. on Automatic Face and Gesture Recognition*, Oct. 1996: pp. 38-44.

[14] D.J. Kang and I.S. Kweon. "A Visual Tracking Algorithm by Integrating Rigid Model and Snakes." *IEEE Conf. on Intelligent Robots and Sys.*, Nov. 1996: 2, pp. 777-784.

[15] T. Kanungo, B. Dom, W. Niblack, and D. Steele. "A Fast Algorithm for MDL-Based Multi-Band Image Segmentation." *IEEE Conf. CVPR*, Jun. 1994: pp. 609-616.

[16] D.E. Knuth. *The Art of Computer Programming.*: Vol 2, pp. 34-54. Addison-Wesley Publishing Company, Inc., 1969.

[17] A. Leonardis and H. Bischof. "Dealing with occlusion in the eigenspace approach." *IEEE Conf. on CVPR*, Jun. 1996: pp. 453-458.

[18] W.Y. Ma and B. S. Manjunath. "Edgeflow: a technique for boundary detection and image segmentation." *IEEE Trans. on Image Proc.*, 2000: 9(8), pp. 1375–1388.

[19] B. North, A. Blake, M. Isard and J. Rittscher, "Learning and Classification of Complex Dynamics," *IEEE Trans. on PAMI*, 2000: 22(9), pp. 1016–1034.

[20] F. Pedersini, A. Sarti, and S. Tubaro. "Accurate Feature Detection and Matching for the Tracking of Calibration Parameters in Multi-Camera Acquisition Systems." *IEEE Conf. on Image Proc.*, Oct. 1998: 2, pp. 598-602.

[21] S. Sclaroff and J. Isidoro. "Active Blobs." *IEEE ICCV*, Jan. 1998: pp. 1146-1153.

[22] J. Shi and C. Tomasi. "Good Features to Track." *IEEE Conf. on CVPR*, Jun. 1994: pp. 593-600.

[23] A. Singh and M. Shneier. "Grey Level Corner Detection: A Generalization and a Robust Real Time Implementation." *CVGIP*, Jul. 1990: 51(1), pp. 54-69.

[24] D. Terzopoulos and K. Fleischer. "Deformable Models." *The Visual Computer*: 4, pp. 306-331.

[25] C. Toklu et al. "Tracking Motion and Intensity Variations Using Hierarchical 2-D Mesh Modeling for Synthetic Object Transfiguration." *Graphical Models and Image Proc.*, November 1996: 58(6): pp. 553-573.

[26] F. de la Torre, S. Gang, and S. McKenna. "View-Based Adaptive Affine Tracking." *Lecture Notes in Computer Science*, 1998: 1406, pp. 828-842.

[27] M. Turk and A. Pentland. "Eigenfaces for Recognition." *J. of Cognitive Neuroscience*, Jan. 1991: 3(1), pp. 71-86.

[28] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry." *INRIA*, May 1994: No 2273.

[29] S.C. Zhu and A. Yuille. "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Mulitband Image Segmentation." *IEEE Trans. on PAMI*, Sep. 1996: 18(9), pp. 884-900.

[30] C. Gentile. "Robust Tracking with Parts in the Presence of Severe Occlusion." Ph.D. Thesis, The Pennsylvania State University, 2001.