Outer-layer based tracking using entropy as a similarity measure

Vincent Garcia, Sylvain Boltz, Eric Debreuve, Michel Barlaud

OUTER-LAYER BASED TRACKING USING ENTROPY AS A SIMILARITY MEASURE

Vincent Garcia, Sylvain Boltz, Éric Debreuve, and Michel Barlaud

Laboratoire I3S, Université de Nice - Sophia Antipolis
2000 route des Lucioles, 06903 Sophia Antipolis, FRANCE
{garciav,boltz,debreuve,barlaud}@i3s.unice.fr

ABSTRACT

Tracking can be achieved using region active contours based on homogeneity models (intensity, motion...). However the model complexity necessary to achieve a given accuracy might be prohibitive. Methods based on salient points may not extract enough of these for reliable motion estimation if the object is too homogeneous. Here we propose to compute the contour deformation based on its neighborhood. Motion estimation is performed at contour samples using a block matching approach. First, partial background masking is applied. Since outliers may then bias the motion estimation, a robust, nonparametric estimation using entropy as a similarity measure between blocks is proposed. Tracking results on synthetic and natural sequences are presented.

***Index Terms*—** Tracking, entropy, block matching, partial background masking

1. INTRODUCTION

The segmentation of video objects is a low level task required for many applications, for example in cinematography. The term "rotoscoping" used in cinematographic post-production corresponds to the all-digital process of tracing outlines over digital film images to produce digital contours in order to allow special visual effects. The segmentation is usually performed manually and frame by frame by so-called animators. As a consequence, it is a long, repetitive, and expensive task. The rotoscoping problem is too complex to define a fully-automatic algorithm. In this paper, we focus on the tracking of an object (*i.e.*, the extraction of the object contour for all frames of the sequence) given an initial, hand-edited contour in the first frame. Some methods based on active contours use global (*i.e.*, region) information [1, 2, 3]. They are usually based on a notion of (possibly non-trivial) homogeneity of the object. Other active contour based methods use motion information [4]. In both cases, if the object is complex or has a complex motion, this description might be difficult to establish and, in the end, not accurate enough to guarantee an accurate tracking.

A local approach [5] proposes to estimate the contour motion from a set of temporal trajectories of keypoints. The resulting tracking is accurate and is robust to occlusions. However, there might not be enough keypoints close to the object contour and, consequently, the tracking may not be accurate enough. In particular, this can happen if the object is rather homogeneous.

In this paper, we propose a tracking method based on the motion estimation of the contour neighborhood. The method is an active contour method where the initial contour is hand-edited in the first frame of the video, and the contour is deformed frame by frame. The contour is discretized in a set of samples according to its representation (polygon, spline, *etc.*). The contour motion is computed by estimating the motion of its samples with a block matching based method. The first contribution of this paper is the use of partial background masking. In image processing, motion estimation, and more generally, parameter estimation, is often based on parametric assumptions (Gaussian, Laplacian, mixture models). However, these assumptions may be false and, consequently, the parameters may not be estimated correctly. The second contribution of this paper is a non-parametric estimation using entropy as a similarity measure between blocks. In practice, entropy is robust to outliers [6]. This property is essential to circumvent the partial aspect of background, otherwise necessary. According to the results on synthetic and natural sequences, the proposed tracking method is accurate.

The paper is organized as follows: Section 2 presents a tracking method based on a matching using partial background masking. Section 3 improves the tracking by using a non-parametric approach. Section 4 shows and discusses some tracking results. Finally, Section 5 concludes.

This work was partly supported by *"Le Conseil Régional Provence-Alpes-Côte d'Azur"*, France.

2. MATCHING BASED ON PARTIAL BACKGROUND MASKING

2.1. Context and classical approach

Let $F_1, \cdots, F_n$ be the frames of a video. Let C_1 be a hand-edited contour in frame F_1 segmenting the object of interest. Assuming that the object contour C_m in frame F_m is known, the problem is to compute the contour C_{m+1} of the object in the next frame. The motion of C_m will be estimated from the (inside) neighborhood of the contour. To account for complex boundary deformation, the contour is discretized and the motion estimation is performed locally at every samples. The samples, moved by their local motion, are then interpolated to form the new contour. The sample motion is assumed to be a translation. This assumption does not restrict the overall motion of the object if C_m is discretized finely enough. In particular, the object can be articulated. The motion of each sample is estimated using a block matching approach [7]: a square block B_i is centered on sample s_i and the block in frame F_{m+1} corresponding to the optimum of a given similarity measure defines the motion v_i of s_i

$$v_i = \arg\min_u \sum_{x \in B_i} \varphi(r_m(x, u)) \tag{1}$$

where φ is a positive function to be defined, and $r_m(x, u)$ is the residual $F_m(x) - F_{m+1}(x + u)$. The function φ must be robust to outliers. This condition excludes choosing the classical sum of squared differences (SSD) criterion $\varphi(r) = r^2$. Functions used in robust estimation [8] could be chosen instead, in particular, $\varphi(r) =$

$|r|$ corresponding to the sum of absolute differences (SAD) criterion. Note that the similarity measure (1) is minimal (and not maximal) when the similarity between the blocks is maximal. For conciseness, the similarity measure will be called criterion in the following. Starting from the classical block matching (1), the different evolutions leading to the proposed method will be justified and illustrated using two synthetic sequences (see Fig. 1) of 300×300 pixels. In these sequences, a textured or homogeneous object is translated horizontally by 4 pixels every frame. The background is fixed. Let us denote these sequences S_{tex} and S_{hom} respectively. The study will focus on two blocks B_l (left block) and B_r (right block) of 33×33 pixels centered around samples s_l and s_r of the object contour (see Fig. 1).

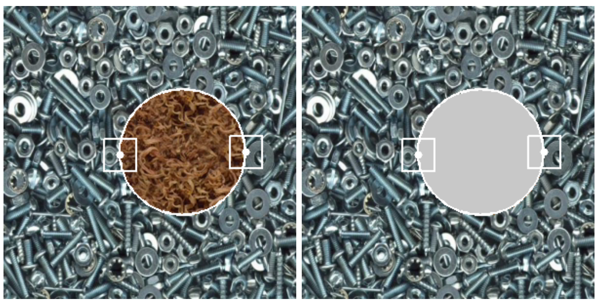
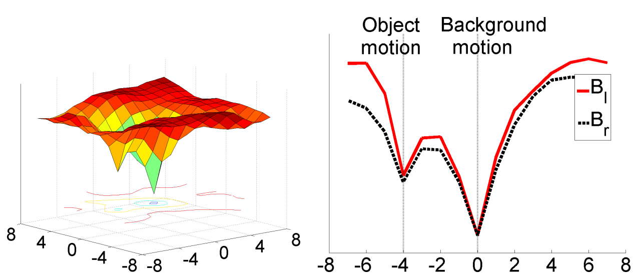


Fig. 1. Object contour C_m in frame F_m of the synthetic sequences S_{tex} (left) and S_{hom} (right). Block B_l (on the left side of the object) and block B_r (on the right side of the object) are respectively centered on samples s_l and s_r.

2.2. Partial background masking



(a) SAD criterion for B_l

(b) 2D profiles of the SAD criteria for B_l (solid line) and B_r (dashed line)

Fig. 2. SAD criterion on a search window of 15×15 pixels for the blocks B_l and B_r in sequence S_{tex}. The profiles of the SAD criterion for B_l and B_r pass through their respective global minimum and are parallel to the X-axis which represents horizontal motion. The profiles were independently scaled to fit in the figure. The object motion estimation fails due to the presence of a majority of background pixels.

As illustrated on Fig. 1, B_l and B_r contain some background. The proportion of background is even greater than the proportion of object in this example. More generally, this observation is true if the object is locally convex at s_l and s_r. Fig. 2(a) shows a plot of the criterion for B_l in sequence S_{tex} using the measure described in (1) with $\varphi(r) = |r|$ over a search window of $[-7, 7] \times [-7, 7]$. Fig. 2(b) shows two profiles corresponding to the criteria computed for B_l and B_r. Due to the large majority of background pixels, the global minimum of the criteria corresponds to the background motion. Actually, this is the correct behavior of a motion estimation method. (Note the presence of a local minimum corresponding to the object motion.) However, the problem here is to assign the object motion to samples. Therefore, we propose to use the domain[1] D_m defined by C_m as a mask. The block truncated with this mask is denoted by $\Omega_i = B_i \cap D_m$. The matching is then performed as follows

$$v_i = \arg\min_u \sum_{x \in \Omega_i} \varphi(r_m(x, u)) \,. \tag{2}$$

Fig. 3 shows that the motion is accurately estimated using truncation

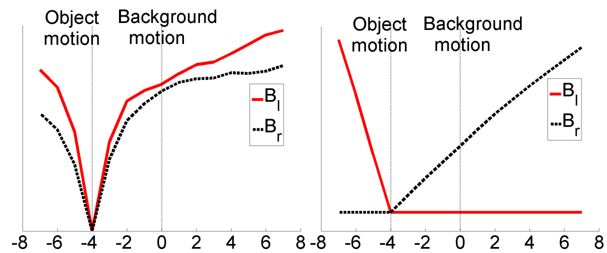


Fig. 3. 2D profiles of the SAD criterion for B_l and B_r in sequences S_{tex} (left) and S_{hom} (right). The profiles were independently scaled to fit in the figure. The motion estimation is accurate with S_{tex} using truncated blocks but it fails with S_{hom}.

for the sequence S_{tex}. However, if the object is relatively homogeneous as in sequence S_{hom}, it might not contain enough structure to allow a reliable motion estimation. The criterion appears flat inside the object domain D_m: the solution of the motion estimation is not unique. In such a case, the boundary can help finding the correct motion by providing the necessary structure. The proposed way of including the object boundary is to dilate D_m before masking B_i. Let d_n be the morphological dilation based on a circular structuring element of radius n. The dilated version of Ω_i is given by

$$\tilde{\Omega}_i = B_i \cap d_n(D_m). \tag{3}$$

Then, the motion is estimated as follows:

$$v_i = \arg\min_u \sum_{x \in \tilde{\Omega}_i} \varphi(r_m(x, u)) \tag{4}$$

Fig. 4 shows the profiles of the SAD criterion for B_l and B_r as a

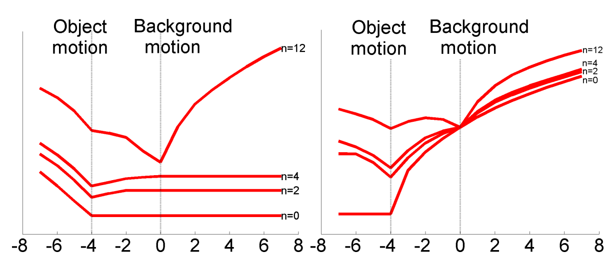


Fig. 4. 2D profiles of the SAD criterion for B_l (left) and B_r (right) for several values of the radius n of the morphological dilation for the sequence S_{hom}.

function of the radius n of the morphological dilation. (Note that

[1] D_m is the domain whose boundary ∂D_m is equal to C_m.

a dilation of 0 pixels corresponds to the truncation presented previously.) It shows that, using the morphological dilation, the object motion is accurately estimated in spite of the homogeneity of the object, at least for a certain range of n. Indeed, if n gets too large, the background becomes dominant and causes the motion estimation to fail as mentioned previously. Consequently, n should belong to an interval $[n_{\min}, n_{\max}]$ where $n_{\min} > 0$. One could look for an *optimal* dilation radius by analysing the object and background textures. Another approach is to choose the radius heuristically while modifying the motion estimation method to enlarge the $[n_{\min}, n_{\max}]$ interval as much as possible by increasing $n_{\max}$. Although SAD or other functions used in robust estimation already ensure that $n_{\max}$ is not too small, we propose to use a non-parametric approach potentially more robust to outliers.

3. MOTION ESTIMATION USING NON-PARAMETRIC ESTIMATION

While solving the problem of a possible lack of structure of a block, the morphological dilation step proposed in Section 2.2 includes outliers (background pixels) in the motion estimation process. Motion estimation using (4) implicitly corresponds to making a parametric assumption on the distribution[2] of the residual r_m. For example, the assumption made is a Gaussian distribution if $\varphi(r) = r^2$ or a Laplacian distribution if $\varphi(r) = |r|$. This assumption is false in general (see below and Fig. 5) and, consequently, the motion may not be estimated correctly. We propose to remove the parametric assumption by letting the motion estimation depends on the true distribution of the residual. Since this distribution is unknown, it will be replaced with an estimation p. The proposed criterion is the Ahmad-Lin approximation of the entropy [9] of the residual

$$v_i = \arg\min_u -\frac{1}{|\tilde{\Omega}_i|} \sum_{x \in \tilde{\Omega}_i} \log(p(r_m(x,u))) \quad (5)$$

where p is obtained using the Parzen method [10]. The choice of this criterion was motivated by the fact that entropy is a measure of dispersion and, ideally, the residual obtained for the true motion is, informally speaking, a “Dirac delta function”, *i.e.*, a distribution with minimal dispersion.

Fig. 5 shows the residual distribution obtained with the true motion of B_l in sequence S_{tex}. The distribution has a main peak and some lower peaks corresponding to the outliers (mismatches of the background pixels). The distribution is clearly not parametric.

Fig. 6 shows the profiles of B_l for sequence S_{tex} for the entropy-based criterion and the SAD criterion for two morphological dilation radii. With both radii, the proportion of object pixels is greater than the proportion of background pixels. The SAD criterion allows to find the correct motion for the radius of 10 pixels but fails with the radius of 14 pixels. The entropy-based criterion allows to estimate the correct motion in both cases, which illustrates its better robustness to outliers (in other words, $n_{\max}$ is larger).

Note that the robustness to outliers is not only required because of the morphological dilation step. Indeed, suppose that this step is removed. Contour C_m cannot be assumed perfect. It probably contains some background. If, consequently, the motion estimation performed on Ω_i is disturbed, the contour in the next frame might contain even more background and the tracking may fail after a few frames.

[2]Parametric in the sense that the distribution is defined by a small set of parameters, *e.g.*, the mean and the variance for a Gaussian distribution.

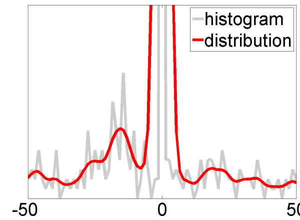


Fig. 5. Close up of the distribution of the residual obtained with the true motion of B_l for the synthetic sequence S_{tex}.

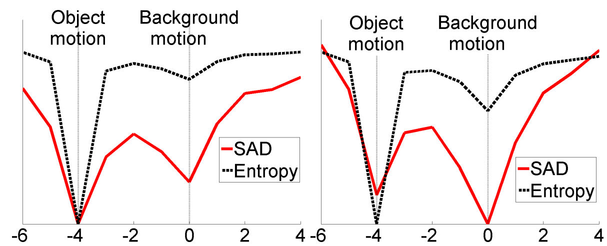


Fig. 6. Comparison of the 2D profiles of the SAD and entropy-based criteria for B_l on sequence S_{tex}. The morphological dilation radius is $n = 10$ pixels for the left figure and $n = 14$ pixels for the right figure. The use of the entropy-based criterion increases the robustness of the motion estimation to outliers.

4. EXPERIMENTS

In this paper, a spline curve is used to represent contour C_m. Actually, C_m is discretized into samples s_i and is interpolated by a spline curve. The motion v_i of s_i is estimated using (5), and C_{m+1} is represented by the spline interpolating samples $s_i + v_i$. However, the proposed method does not depend on the contour representation. Indeed, if the object boundary has sharp edges, contour C_m should probably be represented by a polygon whose vertices are the samples s_i.

Criterion 5 was minimized using a fast, suboptimal search within a search window [11].

Fig. 7. Tracking on sequence *Soccer*. The left figure shows the initial, hand-edited contour C_1 on frame F_1. The right figure shows the computed contour C_5 on frame F_5 (solid line) with C_1 superimposed (dashed line).

The proposed method has been tested on two natural SD (SD=704× 576 pixels) video sequences. Sequence *Soccer* (see Fig. 7) shows a man walking and sequence *Ice* (see Fig. 9) shows a woman skating backward. The motion of these two characters is articulated. The size of block used for the experiments was 33×33 pixels.

Fig. 8 presents the percentage of misclassified pixels (the area of

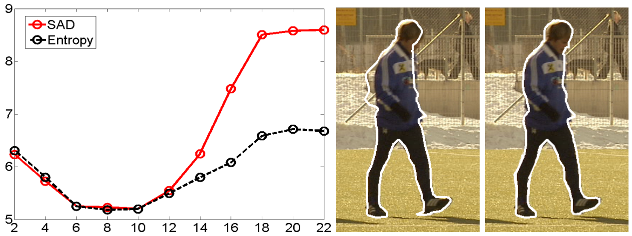


Fig. 8. Comparison of the SAD and entropy-based criteria. The left figure shows the percentage of misclassified pixels (the area of the symmetric difference of the mask of the computed segmentation and the mask of a handmade, *ground truth* segmentation) as a function of the morphological dilation radius for both the parametric (SAD) and the non-parametric (entropy) approaches. Beyond a dilation radius of 22 pixels ($22 \approx \sqrt{2}\frac{\text{blocksize}}{2}$), the motion estimation uses no background masking. The figures on the right show (from left to right) the tracking using SAD and the tracking using the entropy-based criterion on frame F_5.

Fig. 9. Tracking on sequence *Ice*. The left figure shows the initial, hand-edited contour C_1 on frame F_1. The right figure is a close up of the computed contour C_5 on frame F_5 (solid line) with C_1 superimposed (dashed line).

the symmetric difference of the mask of the computed segmentation and the mask of a handmade, *ground truth* segmentation) as a function of morphological dilation radius for both the parametric (SAD) and the non-parametric (entropy) approaches and for 5 frames of the sequence *Soccer*. The ground-truth has been hand-edited frame by frame. Note that a dilation of 22 pixels ($22 \approx \sqrt{2}\frac{\text{blocksize}}{2}$) corresponds to consider no background masking as exposed in Section 2.1.

First, we note that the use of partial background masking allows to decrease the percentage of misclassified pixels in comparison to the use of full background masking or no background masking. More precisely, the decreasing is approximately 40% for the parametric approach and up to 45% for the non-parametric one. Moreover, the optimal radius for both criteria, equal to $n = 10$ pixels, corresponds neither to full background masking nor to no background masking. This indicates that the object is both textured and homogeneous.

Second, we have already shown on synthetic experiments that the use of entropy-based criterion increases the robustness to outliers (background pixels added by the morphological dilation) of the motion estimation. Fig. 8 confirms this behavior on real conditions. Indeed, beyond the global minimum, we note that the increasing of the percentage of misclassified pixels is more important using parametric approach (SAD) than non-parametric approach (entropy). The entropy-based criterion more robust to outliers.

Finally, Fig. 7 and Fig. 9 present the computed contours on frame F_5 with the initial contour superimposed. In both sequences, the characters are accurately tracked in spite of the articulated motion and the presence of homogeneous region. The radius n of the morphological dilation used was 10 pixels.

5. CONCLUSION

We proposed a tracking method based on a contour motion estimation using a local, robust similarity measure. The motion is estimated at contour samples using a block matching approach with partial background masking. This allows to remove most of the outliers (background pixels) while accounting for the object boundary in order to have enough structure in case the object is homogeneous. The motion estimation using a classical criterion may be biased since the partially masked block contains some outliers. In contrast, the use of the proposed nonparametric, entropy-based motion estimator significantly increases the robustness to outliers.

6. REFERENCES

[1] A. Blake and M. Isard, *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*, Springer-Verlag, New-York, NY, USA, 1998.

[2] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, “Image segmentation using active contours: Calculus of variations or shape gradients?,” *SIAM Journal on Applied Mathematics*, vol. 1, no. 2, 2003.

[3] F. Precioso, M. Barlaud, T. Blu, and M. Unser, “Robust real-time segmentation of images and videos using a smooth-spline snake-based algorithm,” *IEEE Transactions on Image Processing*, vol. 7, pp. 910–924, 2005.

[4] D. Cremers and S. Soatto, “Motion competition: A variational framework for piecewise parametric motion segmentation,” *International Journal of Computer Vision*, vol. 62, pp. 249–265, 2005.

[5] V. Garcia, É. Debreuve, and M. Barlaud, “A contour tracking algorithm for rotoscopy,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006.

[6] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, and L. Pronzato, “A minimum-entropy procedure for robust motion estimation,” in *IEEE International Conference on Image Processing (ICIP)*, Atlanta, GA, USA, October 2006.

[7] T. Koga, K. Linuma, A. Hirano, Y. Iijima, and T. Ishiguro, “Motion compensated interframe coding for video conferencing,” in *National Telecommunications Conference (NTC)*, New Orleans, LA, USA, 1981.

[8] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Deterministic edge-preserving regularization in computed imaging.,” *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[9] I. A. Ahmad and P. E. Lin, “A nonparametric estimation of the entropy for absolutely continuous distributions,” *IEEE Transactions on Information Theory*, vol. 36, pp. 688–692, 1989.

[10] E. Parzen, *Stochastic processes*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[11] S. Zhu and K. K. Ma, “A new diamond search algorithm for fast block-matching motion estimation,” *IEEE Transactions On Image Processing*, vol. 9, no. 2, February 2000.