MULTI-FEATURE VECTOR FLOW FOR ACTIVE CONTOUR TRACKING

Joanna Isabelle Olszewska, Christophe De Vleeschouwer and Benoit Macq

Université catholique de Louvain (UCL), Communications Laboratory (TELE)
Place du Levant, 2, B-1348 Louvain-la-Neuve
http://www.tele.ucl.ac.be/

ABSTRACT

In order to achieve both fast tracking and accurate object extraction, we present in this paper an original real-time active contour method, incorporating different feature maps into a common and homogeneous framework, defined by the multi-feature vector flow (MFVF). The MFVF active contour approach does not require any target prior model, and enables precise tracking of mobile deformable objects. The use of the MFVF, resulting from multiple selected features, brings robustness into the system towards complex situations, while our computationally efficient implementation of the MFVF scheme reaches the required speed range for tracking process. The proposed method has been successfully tested on real-world video sequences.

***Index Terms*—** Video Real-Time Tracking, Active Contours, Blobs, Gradient Vector Flow, Feature Combination

1. INTRODUCTION

Many computer vision applications like video-surveillance or human-machine interfaces, necessitate tracking and object extraction as corner stones. Hence, robust, accurate and real-time tracking, in dynamic natural situations, still remains a challenge.

Usual approaches using template matching [1], region growing [2] or target description modeling [3], localize the target-object only by centroid. Moreover, they mostly encounter drawbacks like the lack of flexibility in face of highly changing target shape and appearance. Recent studies tend then to improve their tracking performances by combining several features, as human vision system is processing itself [4], in a sequential [1] or hierarchical way [2]. Nevertheless, these techniques are usually restricted to some specific feature choices and often become time-consuming, therefore, not well adapted to online tracking.

In our work, we have developed a parallel feature fusion scheme, that is computationally effective. The new method integrates, thus at the same level, an extensive number of different feature maps, creating a multi-feature vector flow (MFVF) field. Each feature is chosen depending on the focused application and has an automatically optimized impact on the global result. The MFVF coupled to a snake formalism [5] does not need any prior knowledge on the target appearance, and yields to an efficient real-time tracking system, providing the entire target contour.

The paper is organized as follows. In Section 2, the proposed multi-feature vector flow (MFVF) is described. In Section 3, the MFVF is applied to a real-time active contour method, defining an innovative robust tracking framework. The approach performances are presented on video-surveillance sequences, in Section 4. Finally, conclusions are given in Section 5.

Joanna Olszewska is supported by a FRIA grant of the Belgian NSF.

2. MULTI-FEATURE VECTOR FLOW (MFVF)

The Multi-Feature Vector Flow (MFVF) defines a new active contour external force Ξ based not only on the image intensity gradient [6], [7], [8] but on the gradient of any feature map.

In our approach, at first, feature maps have to be generated from the selected features. Then, the feature vector flows are computed, in parallel, using an isotropic diffusion process, like described in Section 2.1. Finally, the MFVF is obtained by a fusion of the feature vector flow fields, exposed in Section 2.2.

2.1. Feature Vector Flow Field

The MFVF field $\Xi(x, y)$ is defined as a combination of N_F feature vector fields $\Xi_j(x, y) = [\xi_{uj}(x, y), \xi_{vj}(x, y)]$. Each vector Ξ_j minimizes the following functionality ε_j

$$\varepsilon_j = \iint \mu_j(\xi_{uxj}^2 + \xi_{uyj}^2 + \xi_{vxj}^2 + \xi_{vyj}^2) + (f_{xj}^2 + f_{yj}^2)((\xi_{uj} - f_{xj})^2 + (\xi_{vj} - f_{yj})^2)\, dxdy \tag{1}$$

where the diffusion parameter μ_j could be set according to the amount of image noise and f_j is a feature map computed from the corresponding j^{th} adapted feature.

Based on these formulations, the j^{th} feature vector flow field Ξ_j can be found by solving the Euler equations

$$\mu_j \nabla^2 \xi_{uj} - (\xi_{uj} - f_{xj})(f_{xj}^2 + f_{yj}^2) = 0 \tag{2}$$

$$\mu_j \nabla^2 \xi_{vj} - (\xi_{vj} - f_{yj})(f_{xj}^2 + f_{yj}^2) = 0 \tag{3}$$

where ∇^2 is the Laplacian operator.

Considering ξ_{uj} and ξ_{vj} as functions of time, leads to Equations 4 and 5

$$\xi_{ujt}(x, y, t) = \mu_j \nabla^2 \xi_{uj}(x, y, t) - [\xi_{uj}(x, y, t) - f_{xj}(x, y)].[f_{xj}(x, y)^2 + f_{yj}(x, y)^2] \tag{4}$$

$$\xi_{vjt}(x, y, t) = \mu_j \nabla^2 \xi_{vj}(x, y, t) - [\xi_{vj}(x, y, t) - f_{yj}(x, y)].[f_{xj}(x, y)^2 + f_{yj}(x, y)^2]. \tag{5}$$

The Equations 4 and 5 also known as generalized diffusion equations, are then decoupled. Hence, these scalar partial differential equations in ξ_{uj} and ξ_{vj}, can be computed in a separate way. The solution of the j^{th} feature vector flow Ξ_j is therefore obtained after discretization and iteration of Equations 4 and 5.

2.2. Feature Vector Flow Fusion

In order to treat the different types of low and high-level features in a homogeneous framework, and to process in a computationally efficient way, the multi-feature vector flow MFVF fusion is formalized by taking a linear model given by a weighted sum:

$$\Xi(x,y) = \sum_{j=1}^{N_F} w_j \; \Xi_j(x,y). \tag{6}$$

The weight values w_j are depending on the ability of each feature to localize accurately the object-target in a given image. In this sense, we first introduce, in Section 2.2.1, a new criterion called Segmentation Method Attenuation (SMA) to evaluate the segmentation quality resulting of each feature, in a useful way for our approach. Then, we describe, in Section 2.2.2, our weight computation algorithm.

2.2.1. *Segmentation Method Attenuation (SMA)*

The segmentation method attenuation (SMA) is a new criterion providing the evaluation of shape similarity and object location concordance between the reference mask area and the extracted one.

Considering a reference image I_r, we have defined this shape similarity criterion as the attenuation introduced by the object extraction method, expressed in dB,

$$SMA(M_r, R) = 10 \log \frac{\sum\limits_{(x_i,y_i)\in I_r} (M_r(x_i,y_i))^2}{\sum\limits_{(x_i,y_i)\in I_r} (R(x_i,y_i))^2} \tag{7}$$

with R, the overlapped region defined as the intersection of the reference mask $M_r(x_i, y_i)$ and the extracted mask $M_e(x_i, y_i)$,

$$\forall (x_i,y_i) \in I_r,$$
$$R(x_i,y_i) = \begin{cases} 1 & \text{if } M_r(x_i,y_i) + M_e(x_i,y_i) = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The masks M are simply defined as

$$\forall (x_i,y_i) \in I,$$
$$M(x_i,y_i) = \begin{cases} 1 & \text{if } (x_i,y_i) \in \text{video object}, \\ 0 & \text{if } (x_i,y_i) \in \text{background}. \end{cases} \tag{9}$$

In the best case, the $SMA(M_r, R)$ is equal to zero when the extracted object region is corresponding to the reference one, considering its area location. The worst situation occurs when there is no overlap between the extracted object and the reference one, leading to a total attenuation, therefore an infinite SMA value. In between, the SMA value is varying, related to the match between the reference mask and the extracted one. Thereby, the SMA provides, in the sense of [9] recommendations, a meaningful measure that could be rapidly evaluated and applied to score a large set of feature maps.

2.2.2. *Weight Computation*

The feature weights w_j are computed in two phases. Firstly, the weight are fixed to an equal value, mentioned in Equation 10. The Equation 6 is solved for these values, providing an initial MFVF. Based on these results, a reference mask M_r is defined, by the mean of a MFVF active contour as described in Section 3.2. Secondly, respecting the condition $\sum_j w_j = 1,\ for\ j = 1...N_F$, the weight values are computed thanks to Equations 11 and 12. The SMA_j is the segmentation method attenuation (see Section 2.2.1), computed for the j^{th} extracted feature which defines the corresponding M_e.

- Initialization:

$$w_{j_0} = \frac{1}{N_F} \qquad j = 1...N_F \tag{10}$$

- Update:

$$w_j = \frac{g_j}{\sum\limits_{j=1}^{N_F} g_j} \tag{11}$$

The coefficient g_j is defined as follows,

$$\text{if} \sum_{j=1}^{N_F} SMA_j \neq 0,$$
$$g_j = \begin{cases} (SMA)_j^{-1} & \text{if } SMA_j \neq 0 \\ k \cdot max\{g_{j' \neq j}\} & \text{if } SMA_j = 0, \text{with } k \in \mathbb{Z}_0^+ \end{cases}$$
$$\text{otherwise,} \quad g_j = \frac{1}{N_F}. \tag{12}$$

Finally, the feature weight values found in Equation 11, are replaced into Equation 6, to compute the resulting MFVF field.

The proposed multi-feature vector flow fusion scheme has the main advantage to offer the ability to directly control the impact of each feature on segmentation accuracy as well as on tracking quality. Moreover, our framework enables the grouping of different nature image-information, like edge or color, in one vectorial field (MFVF), that could be used as an external force in the active contour mechanism, described in Section 3.

3. MFVF ACTIVE CONTOUR

We present in this section a fast and robust active contour method based on the MFVF. Our approach leads to a generic multi-feature tracker and relies on two major conception steps. At first, the selection of the features, in order to generate the MFVF, is explained in Section 3.1. Then, the model implementation of the active contour itself, converging thanks to the MFVF external force, is described in Section 3.2. Furthermore, all the system is designed to be compatible with real-time tracking as shown in Section 4.

3.1. Selected Features

In this paper, two different features are taken to illustrate our MFVF mechanism : the edges and the blobs. The choice is based on their complementary structural properties as well as their different extracted content. Indeed, the edges define the object border shape and are computed thanks to image intensity gradient as mentioned in Section 3.1.1. On the other hand, the blobs provide both the region of interest containing the target-object, and information on its appearance, by background subtraction, described in Section 3.1.2.

(a) (b) (c)

Fig. 1. MFVF fields performances on frames of a video sequence (first row : frame 849; second row : frame 2316), (a) Ξ_1 based on edges, (b) Ξ_2 based on a blob, (c) MFVF resulting field Ξ.

3.1.1. *Edges*

Let $I(x,y) : [0, L_x] \times [0, L_y] \rightarrow \mathbb{R}^+$ be a given image and L_x, L_y, its spatial dimensions. The edges are defined to have strong magnitudes at the boundaries of the objects and are identified, in our case, by the location of zero-crossings of second order derivative operations in the image $I(x,y)$. The edge map f_1 is thus estimated by convolving the image I with the Laplacian of the Gaussian, also known as LoG operator,

$$f_1(x,y) = |\nabla(G_\sigma(x,y) * I(x,y)|^2 \quad (13)$$

where G_σ is a two-dimensional Gaussian function with standard deviation σ.

The feature vector flow field $\Xi_1(x,y)$, shown in Figure 1 (a), is calculated by incorporating Equation 13 into Section 2.

3.1.2. *Blobs*

The blobs are defined by labeled connected regions, using background subtraction. This technique consists in computing the difference between the current image $I(x,y)$ and a background model, and afterwards, in extracting the foreground.

We adopt the Running Gaussian Average for modeling the background, characterized by the mean μ_b and variance σ_b^2, since it is a method well-suited for real-time tracking [10].

The foreground is then determined by Equation 14.

$$F_2(x,y) = \begin{cases} 1 & \text{if } |I(x,y) - \mu_b| > n \cdot \sigma_b, \text{ with } n \in \mathbb{N}_0, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Finally, the morphological operations are applied to the extracted foreground F_2, in order to exploit the existing information on the neighboring pixels,

$$f_2(x,y) = Morph(F_2(x,y)) \quad (15)$$

where the blob map f_2 is thus adapted to our feature vector flow process. This leads to the feature vector flow field $\Xi_2(x,y)$, illustrated in Figure 1 (b).

3.2. MFVF Snake

For computational efficiency, we model the MFVF active contour with a parametric planar curve $\boldsymbol{\mathcal{C}}(s) : [0,1] \rightarrow \mathbb{R}^2$, also called snake [5], represented by a B-Spline formalism.

The active contour evolution, from an initial position towards desired image features, is driven by the dynamic Equation 16, that is depending on internal force, described by contour mechanical properties (α: elasticity, β: rigidity), and on external force Ξ resulting from the image selected features.

$$\boldsymbol{\mathcal{C}}_t(s,t) = \alpha\, \boldsymbol{\mathcal{C}}_{ss}(s,t) - \beta\, \boldsymbol{\mathcal{C}}_{ssss}(s,t) + \Xi \quad (16)$$

(a) Frame 849 (left), zoom on the target (right)   (b) Frame 2316 (left), zoom on the target (right)

Fig. 2. MFVF Tracking results on video-surveillance sequence of 2400 frames.

The MFVF external force Ξ owns the ability of a a large capture range and bi-directional convergence. Its additional capabilities are related to the extracted feature properties.

Therefore, the Equation 16 defines a general multi-feature active contour framework, enabling the use of an extendable number of different object shape and appearance characteristics.

4. TRACKING RESULTS

The presented MFVF active contour approach has been validated on real-world video-surveillance sequences, coming from the CAVIAR standard dataset:

`http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`

Our method implementation allows us to achieve accurate and real-time tracking performances, as shown in Figure 2. Indeed, the fast B-Spline formalism is coupled with an effective computation of the MFVF field, using the principle of [8]. Moreover, the feature maps themselves are efficiently extracted as explained in Section 3.1. The contour initialization, at each frame, is automatically done by means of an enhanced bounding box, centered on the detected blob in the current frame.

The robustness of the MFVF active contour system is provided by the complementary properties of the selected features, whatever the extractor precision. Thus, compare to the fields built only on one feature like illustrated in Figure 1 (a) and (b), the MFVF field owns a more pertinent content in term of vectorial information and conducts to more accurate field results as shown in Figure 1 (c). Hence, the MFVF mechanism property allows the tracker to be robust even in complex situations, like real-world background cluttered with disturbing patterns and similar distracting objects (see Figure 2).

5. CONCLUSIONS AND PERSPECTIVES

This paper presents a new multi-feature vector flow (MFVF) field, enabling an efficient combination of various types of features in a homogeneous way. As shown in this work, the use of the proposed MFVF into the active contour approach, leads to a real-time robust generic framework that is well-suited for accurate tracking of mobile deformable objects in video sequences.

6. REFERENCES

[1] J. Sung and D. Kim, "Large Motion Object Trakcing Using Active Contour Combined Active Appearance Model," in *Proceedings of the IEEE International Conference on Computer Vision Systems*, January 2006, pp. 31–39.

[2] T.-H. Tsai and C.-Y. Lin, "Video Segmentation Using Multiscale Feature Extraction and Fusion," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 2005, pp. IV.3223–IV.3226.

[3] J. Kang, I. Cohen, and G. Medioni, "Object Reacquisition using invariant appearance model," in *Proceedings of the IEEE International Conference on Pattern Recognition*, August 2004, pp. IV.759–IV.762.

[4] C. Siagian and L. Itti, "Rapid Biologically-Inspired Scene Classification Using Features Shared With Visual Attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300–312, February 2007.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, January 1988.

[6] C. Xu and J.L. Prince, "Snakes, shapes and Gradient Vector Flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, March 1998.

[7] H. Yu and C.-S. Chua, "GVF-Based Anisotropic Diffusion Models," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1517–1524, June 2006.

[8] J.I. Olszewska, C. De Vleeschouwer, and B. Macq, "Speeded Up Gradient Vector Flow B-Spline Active Contours for Robust and Real-Time Tracking," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2007, pp. I.905–I.908.

[9] D.P. Young and J.M. Ferryman, "PETS Metrics: On-Line Performance Evaluation Service," in *Proceedings of the IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, October 2005, pp. 317–324.

[10] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, July 1997.