Statistical Prior Based Deformable Models for People Detection and Tracking

Amira Soudani(✉) and Ezzeddine Zagrouba

Equipe de Recherche SIIVA, Laboratoire RIADI, Institut Supérieur d'Informatique, Université de Tunis El Manar, 2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia
amira.soudani@gmail.com, ezzeddine.zagrouba@fsm.rnu.tn

Abstract. This paper presents a new approach to segment and track people in video. The basic idea is the use of deformable model with incorporation of statistical prior. We propose an hybrid energy model that incorporates a global and a statistical based energy terms in order to improve the tracking task even under occlusion conditions. Target models are initialized at the first frame, then predictions are constructed based on motion vectors. Therefore, we apply an hybrid active contour model in order to segment tracked people. Experiments show the ability of the proposed algorithm to detect, segment and track people well.

Keywords: Tracking · Segmentation · Deformable models · Multiple targets · Active contours · Occlusion

1 Introduction

Object tracking is the process of locating moving targets throughout the frames of video sequences. In this aim several works were proposed and divided into three categories [1] based on the tracked feature: Point, Kernel and Silhouette based tracking methods. In this paper, we propose to classify methods of the state of art based on the representation of the tracked targets. In fact, a target object is represented by a model that includes information about its shape and appearance. The shape of the object of interest can be approximated by a basic, articulated or deformable representation. In [2], authors propose a tracking method based on basic representation and particle filtering. Indeed, each target is initialized in the first frame then it is divided into seven parts. Multi-part RGB kernel histograms are computed for each part. The particle filter is used to look for candidates in the current image based on their previous states. For each particle provided by the particle filtering process, a multi-parts representation is computed as same as the models of the observations. Thereafter, they choose the optimal candidate through l_1 regularized least square approach and the appearance model is updated. A tracking method based on articulated representation is presented by [3]. They propose a method for detecting articulated people and estimating their pose from static images based on a new representation of deformable part models. Several tracking methods based on deformable

S. Arik et al. (Eds.): ICONIP 2015, Part III, LNCS 9491, pp. 392–401, 2015.
DOI: 10.1007/978-3-319-26555-1_44

representation are proposed. Authors in [4] use the active contour to model the human body and propose a novel method for tracking the body's contour by combining the colour and the depth cues adaptively. Their method consists on two mainly stages. First, they evolve the active contour to the object's boundary by integrating the edge and the region cues of the depth image and the region cue of the colour image in level set framework. In the second stage, they refine the tracking result provided by the level set method using the two properties of the body surface in depth image. In [5], authors propose an approach to segment and track multiple persons in a video sequence via graph-cuts optimization technique. They extract initial silhouettes that will be modeled by ellipses. Then, a prediction step based on optical flow vectors is used to detect if an occlusion will handle. Hence, they identify the occluding persons by the use of the Chi-squared similarity metric based on the intensity histogram and they update the objects models of the interacting persons. Finally, a segmentation based on graph-cuts optimization is performed based on the predicted models.

In this paper, we address the problem of tracking people based on deformable models. In fact, we focus on region based active contour methods in order to segment and track people. The novelty of our approach is the incorporation of a color constraint in the energy functional which is composed of a local and global region based energy terms. The rest of the paper is organized as follows. Section 2 presents the proposed approach. Section 3 outlines the algorithm followed by experimental results in Sect. 4. Finally, conclusion and perspectives are presented in Sect. 5.

2 Hybrid Active Contour Based Person Segmentation and Tracking

2.1 Shape Description and Initialization

At first, we proceed to a background substraction using a reference frame in order to extract N initial targets to be tracked in the video sequence. We associate to each target $P^{(i,t_0)}, i \in \{1 \dots N\}$ a closed contour $\zeta^{(i,t_0)}$ and its corresponding convex hull $\Psi^{(i,t_0)}$ (Fig. 1).

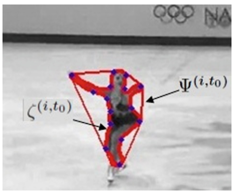


Fig. 1. Shape initialization

2.2 Contour Prediction

Once the people are detected, we will compute a prediction of the contour $\zeta_{pred}^{(i,t)}$ for each object i based on the average of optical flow vectors $d^{(i,t-1)}$ of its previous displacements. Each predicted contour $\zeta_{pred}^{(i,t)}$ is represented by a corresponding convex hull $\Psi_{pred}^{(i,t)}$. These predictions will be useful in the tracking process. In fact, they will be incorporated in the energy functional in order to deduce the final contour of each tracked person.

$$\zeta_{pred}^{(i,t)} = \{x + d^{(i,t-1)} | \, x \in \zeta^{(i,t-1)}\}. \tag{1}$$

2.3 Active Contour Model with Statistical Prior

In this paper, we propose an energy functional based on statistical prior. In fact, for each target i at time t, we will minimize an energy function as follows:

$$E(\zeta^{(i,t)}) = \alpha E_{glob}(\zeta^{(i,t)}) + (1-\alpha) E_{stat}(\zeta^{(i,t)}). \tag{2}$$

Where α is positive user fixed constants. Therefore we add a regularization term in order to keep the curve smooth Eq. 3.

$$E(\zeta^{(i,t)}) = \alpha E_{glob}(\zeta^{(i,t)}) + (1-\alpha) E_{stat}(\zeta^{(i,t)}) + \gamma |\zeta^{(i,t)}|. \tag{3}$$

Where γ is a positive user fixed constant and $|\zeta^{(i,t)}|$ is the contour's length.

Global Energy. Let's denote by $I : \Omega \rightarrow \Re$ a given image function where $I(x)$ is the intensity of pixel $x \in \Omega$. The first term E_{glob} is based on the uniform modeling energy proposed by Chan and Vese [6]. The energy is based on the means intensities inside and outside a closed contour ζ.

$$E_{glob}(\zeta^{(i,t)}) = \int_{\Omega_{in}} (I(x) - u_{in})^2 dx + \int_{\Omega_{out}} (I(x) - u_{out})^2 dx \tag{4}$$

Where u_{in} and u_{out} are respectively the means intensities inside and outside $\zeta^{(i,t)}$, Ω_{in} and Ω_{out} are respectively the areas inside and outside $\zeta^{(i,t)}$. The minimization problem is solved by the use of level set representation. Therefore, the closed contour $\zeta^{(i,t)}$ associated to target i at time t is represented as the zero level set of the signed function $\Phi^{(i,t)}$ such that:

$$\begin{cases} \zeta^{(i,t)} = \{x \in \Omega | \Phi^{(i,t)}(x) = 0\}; \\ inside(\zeta^{(i,t)}) = \{x \in \Omega | \Phi^{(i,t)}(x) > 0\}; \\ outside(\zeta^{(i,t)}) = \{x \in \Omega | \Phi^{(i,t)}(x) < 0\}. \end{cases}$$

Thus the energy functional is reformulated in terms of level set function:

$$E_{glob}(\Phi^{(i,t)}) = \int_{\Omega} H(\Phi^{(i,t)}(x))(I(x) - u_{in})^2 dx + \int_{\Omega} (1 - H(\Phi^{(i,t)}(x)))(I(x) - u_{out})^2 dx. \tag{5}$$

Where $H(\Phi^{(i,t)}(x))$ is an approximation of the Heaviside function whose value is 1 if $\Phi^{(i,t)}(x)$ is positive and null otherwise while u_{in} and u_{out} are defined as follows:

$$u_{in} = \frac{\int_\Omega I(x)H(\Phi^{(i,t)}(x))dx}{\int_\Omega H(\Phi^{(i,t)}(x))dx}, u_{out} = \frac{\int_\Omega I(x)(1 - H(\Phi^{(i,t)}(x)))dx}{\int_\Omega (1 - H(\Phi^{(i,t)}(x)))dx}. \quad (6)$$

Statistical Constraint. In order to improve the tracking task, we will incorporate a statistical prior to constrain the segmentation. We compute the color distribution associated to the background and the foreground which is a gaussian mixture model adjusted respectively to the set of pixels outside and inside the contour $\zeta^{(i,t)}$ of target i at time t. Hence, the statistical term is defined as follows:

$$E_{stat}(\zeta^{(i,t)}) = -\int_{\Omega_{in}} logP(I(x)|\psi_{in})dx - \int_{\Omega_{out}} logP(I(x)|\psi_{out})dx \quad (7)$$

where $P(I(x)|.)$ denotes the Generalized Gaussian distributions assumed to represent the likelihood that a pixel of intensity $I(x)$ belongs to Ω_{in} or Ω_{out} while ψ_{in} and ψ_{out} are the parameters of the distribution respectively inside and outside $\zeta^{(i,t)}$.
Then, we rewrite E_{stat} in terms of level set function

$$E_{stat}(\Phi^{(i,t)}) = -\int_\Omega H(\Phi^{(i,t)}(x))\, logP(I(x)|\psi_{in})\, dx \\ - \int_\Omega (1 - H(\Phi^{(i,t)}(x)))\, logP(I(x)|\psi_{out})\, dx. \quad (8)$$

Final Model

$$E(\Phi^{(i,t)}) = \alpha E_{glob}(\Phi^{(i,t)}) + (1-\alpha)E_{stat}(\Phi^{(i,t)}) + \\ \gamma \int_\Omega \delta(\Phi^{(i,t)}(x))|\nabla\Phi^{(i,t)}(x)|dx. \quad (9)$$

We use the Euler-Lagrange equations to solve the minimization problem. Then the level set function can be updated by gradient descent method and the evolution equation is expressed by

$$\frac{\partial\Phi^{(i,t)}}{\partial t}(x) = \alpha\delta(\Phi^{(i,t)}(x))\left((I(x) - u_{in})^2 - (I(x) - u_{out})^2\right) + \\ (1-\alpha)\,\delta(\Phi^{(i,t)}(x))\left(-\frac{log(p(I(x|\psi_{in})))}{log(p(I(x|\psi_{out})))}\right) + \gamma\,\delta(\Phi^{(i,t)}(x))\left(div\left(\frac{\nabla\Phi^{(i,t)}(x)}{|\nabla\Phi^{(i,t)}(x)|}\right)\right). \quad (10)$$

2.4 Occlusion Handling

A major challenge of tracking multiple objets is the occlusion handling. We inspired form the work of Brox and Weickert [7] who proposed an algorithm for multiple region segmentation based on the idea of competing regions. We consider that the evolution equation is composed of a retreat and an advance term denoted U and V respectively. The retreat term is always negative and aims to move the curve inward whereas the advance component is always positive and tries to move the curve outward along its normal. Then, our goal is to allow multiple contours to compete with each other. To do this, we retain the notion of competition between advance and retreat forces and combine them in a different way. These two terms can be expressed for each tracked object i at time t as follows (Fig. 2):

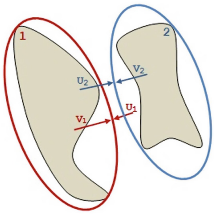


Fig. 2. Advance and retreat components

$$U^{i,t} = \int_{\Omega_{out}} (I(x) - u_{out})^2 dx - \int_{\Omega_{out}} logP(I(x)|\psi_{out})dx + \frac{\gamma}{2}|\zeta^{(i,t)}|. \tag{11}$$

$$V^{i,t} = \int_{\Omega_{in}} (I(x) - u_{in})^2 dx - \int_{\Omega_{in}} logP(I(x)|\psi_{in})dx + \frac{\gamma}{2}|\zeta^{(i,t)}|. \tag{12}$$

Given two adjacent objects $\{i, j \in 1..N \mid i \neq j\}$, then the corresponding U and V terms are defined:

$$\begin{cases} if\, U^{i,t} < -V^{j,t} then\ U^{i,t} = 0; \\ if\, V^{i,t} < -U^{j,t}\, then\ V^{i,t} = 0. \end{cases}$$

This allows us to better deal with occlusion cases. In fact, this allows a contour to not converge to the adjacent region.

3 Algorithm

The proposed algorithm can be summarized as follow:

Algorithm 1. Overview of the proposed method

1.. Initialization of targets
2. Process at time $t > t_0$
   (a) Compute Optical flow vectors
   (b) Predict Targets
   (c) Energy minimization eq. 10
   (d) Update predictions
   set t=t+1 and return to 2

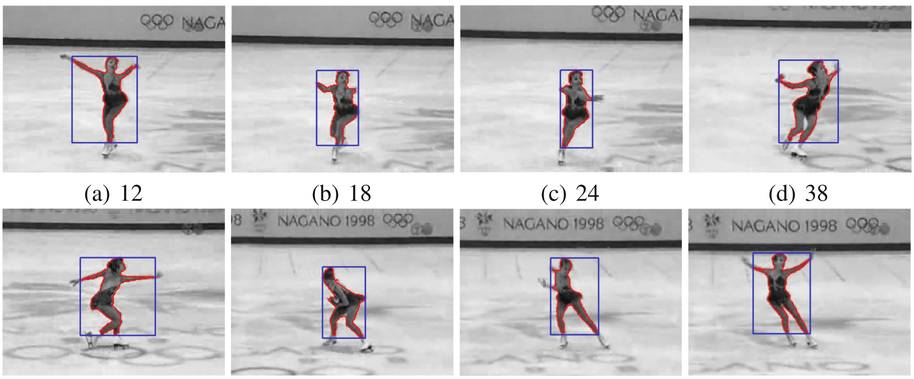
---

4 Experiments

The main goal behind the proposed algorithm is to track persons in video sequence. In order to evaluate the effectiveness and performance of the proposed method, we will test it on several datasets for single and multiple objects tracking (Skater, PETS 2009). Experiments are implemented under Matlab 7.4 in a personal computer with a processor Intel Core i5-3210M, 250 GHZ, 600 GB RAM, Windows 7. The parameters are set as follows: $\alpha = 0.5$, $\beta = 0.5$, $\gamma = 0.5$ and the number of components of the GMM components is set to 5.

4.1 Qualitative Evaluation

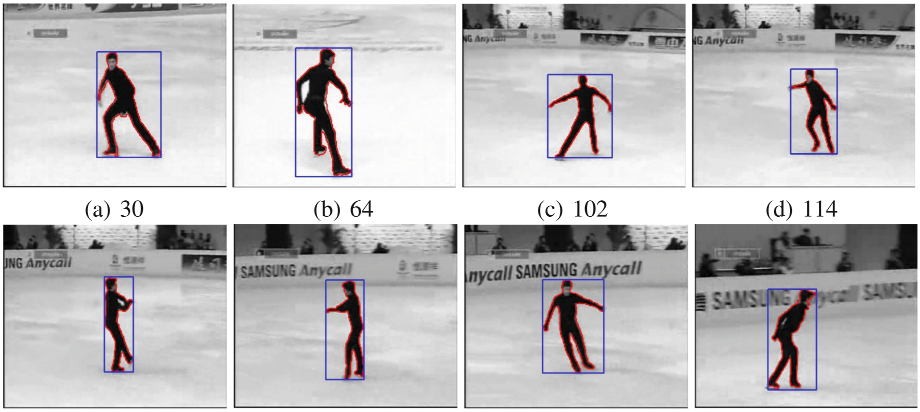
Tracking Single Object. First, we test our method on video sequences including a single object. We present on Figs. 3 and 4 results of the proposed method



(a) 12 (b) 18 (c) 24 (d) 38

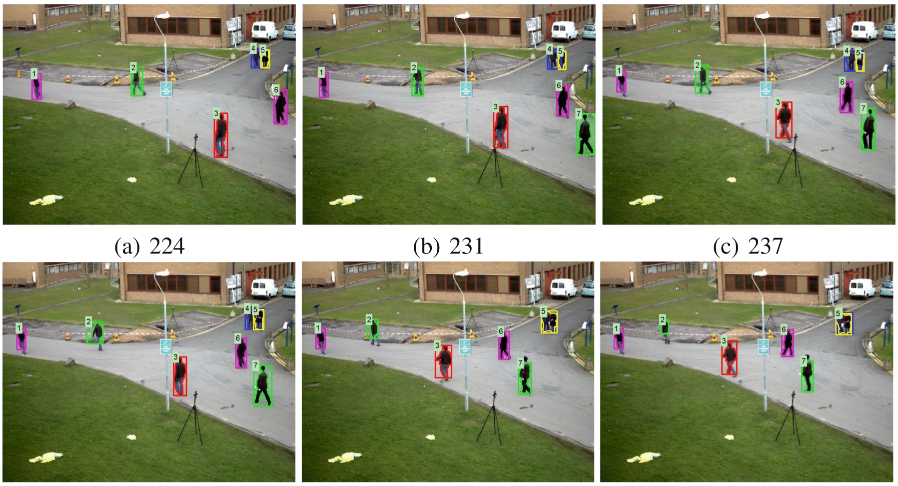
(e) 54 (f) 76 (g) 84 (h) 92

Fig. 3. Female Skater results



(a) 30 (b) 64 (c) 102 (d) 114
(e) 124 (f) 132 (g) 144 (h) 158

Fig. 4. Male Skater results

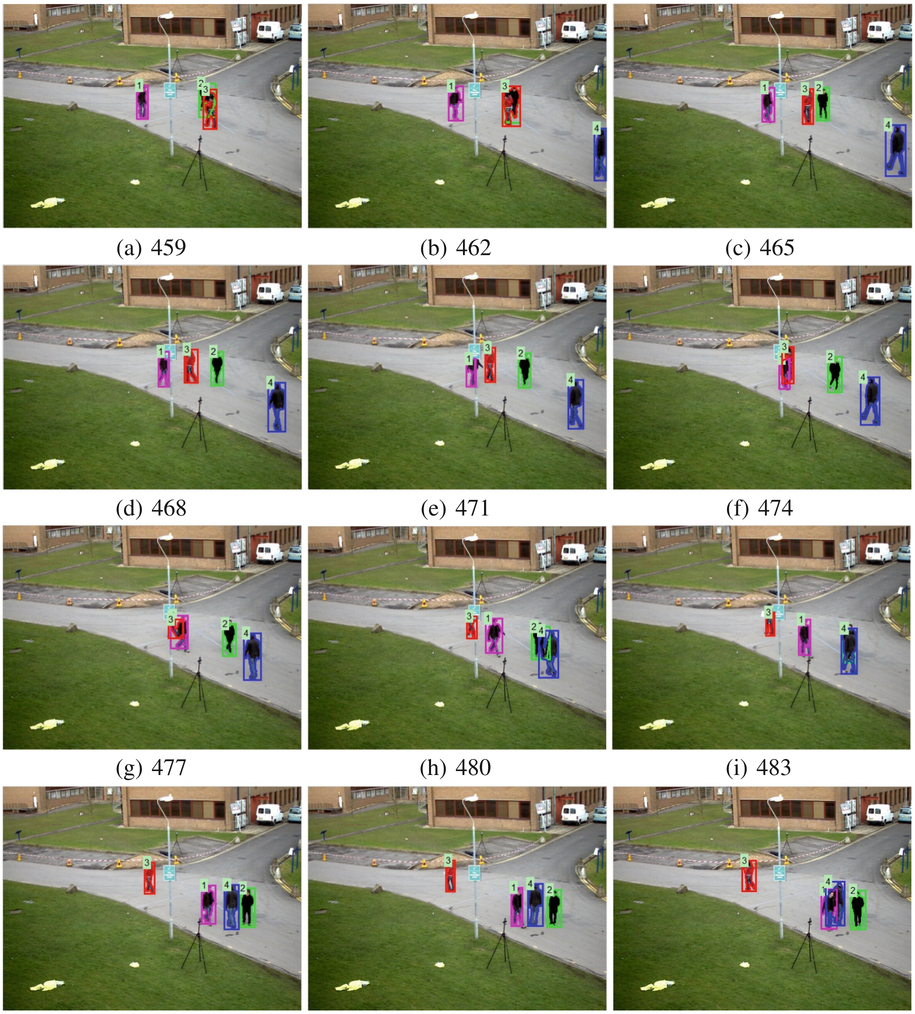


(a) 224 (b) 231 (c) 237
(d) 240 (e) 255 (f) 261

Fig. 5. Results on PETS 2009 sequence

on single object sequences. The obtained results show good contour detection and people tracking. In fact, the contours of the objects are well detected despite the shape deformations and fast motion of the tracked targets.

Tracking Multiple Objects. In this paragraph, we show results of the proposed method on PETS2009 dataset (View 001) (Figs. 5 and 6). People on the sequence are well detected and tracked. We notice the entrance of a new pedestrian in Fig. 5(b) and the proposed method succeeds to detect and to track it along the frames. In Fig. 6, we show occlusion cases in Fig. 6(a, f, h). We deduce



(a) 459 (b) 462 (c) 465
(d) 468 (e) 471 (f) 474
(g) 477 (h) 480 (i) 483
(j) 486 (k) 489 (l) 492

Fig. 6. Results on PETS 2009 sequence

that the proposed algorithm succeeds to detect and track pedestrians even under severe occlusion cases. On the other hand, in Fig. 6(b), a new target appears on the sequence and is well detected and tracked.

4.2 Quantitative Evaluation

We use a selection of metrics proposed in the Video Analysis and Content Extraction (VACE) protocol [8]. In Table 1, we evaluate the overall accuracy (MOTA) and precision (MOTP) of the tracking algorithm. We compare the tracking results on PETS 2009 sequence (View 001) of the proposed method with the works of [5,9,10]. As shown on Table 1, the proposed method gives better tracking and accuracy scores.

Table 1. Quantitative results on the PETS 2009 database (S2.L1).

	LP [9]	SDP [9]	[10]	[5]	Proposed
MOTA	0.82	0.11	0.89	0.67	0.904
MOTP	0.56	0.1	0.562	0.61	0.605
MODA	0.85	0.11	0.908	0.7	0.921
MODP	0.57	0.12	0.573	0.57	0.603

5 Conclusion and Outlines

In this paper we presented a multi-persons tracking approach based on deformable models. We incorporate a statistical prior in the energy term. Experiments were performed on several video sequences including single and multiple objects and provided us good tracking results. As future directions, we aim to improve the segmentation task by incorporating graph-cuts optimization.

References

1. Yilmaz, A., Li, X., Shah, M.: Contour based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 1531–1536 (2004)
2. Shao, J., Dong, N., Tong, M.: Multi-part sparse representation in random crowded scenes tracking. Pattern Recogn. Lett. **34**(7), 780–788 (2013)
3. Yang, Y., Ramanan, D.: Articulated human detection with flexible mixtures of parts. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 2878–2890 (2013)
4. Xu, Y., Ye, M., Tian, Z., Zhang, X.: Locally adaptive combining colour and depth for human body contour tracking using level set method. IET Comput. Vision **8**(4), 316–328 (2014)
5. Soudani, A., Zagrouba, E.: People tracking based on predictions and graph-cuts segmentation. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Li, B., Porikli, F., Zordan, V., Klosowski, J., Coquillart, S., Luo, X., Chen, M., Gotz, D. (eds.) ISVC 2013, Part II. LNCS, vol. 8034, pp. 158–167. Springer, Heidelberg (2013)
6. Chan, T.F., Vese, L.A.: Active contours without edges. Trans. Img. Proc. **10**(2), 266–277 (2001)
7. Brox, T., Weickert, J.: Level set segmentation with multiple regions. IEEE Trans. Image Process. **15**(10), 3213–3218 (2006)
8. Kasturi, R., et al.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 319–336 (2009)

9. Berclaz, J., Fleuret, F., Fua, P.: Multiple object tracking using flow linear programming. In: The 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 1–8 (2009)
10. Andriyenko, A., Schindler, K., Roth, S.: Discrete-continuous optimization for multi-target tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1926–1933 (2012)