# Target and Environments Complexity Characterization for Automatic Visual Tracker Selection in Mobile Robotic Tasks

Antonio Marín-Hernández Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Sebastián Camacho No. 5, Centro, 91000, Xalapa, Mexico. anmarin@uv.mx

#### Abstract

Visual tracking has become a very important task for autonomous mobile robots. In order to accomplish complex tasks a robot needs to track different objects in very different scenarios. Many visual tracking methods have been proposed in literature. However, the lack of robustness to deal with all possible targets and environments conditions makes necessary the implementation on board of more than one method. In this paper, we propose a technique to select from a set of visual trackers the best adapted to a given task. Visual trackers are selected based on analysis and characterization of targets and environments. The characterization is achieved measuring the region complexity over three predefined zones associated to the target. Four visual tracking methods have been used: a) active contours b) 1D correlation c) set of points tracker and d) template differences.

# 1. Introduction

Visual tracking has become a very important task with a wide spectrum of applications, including people tracking in videoconferencing, clinical performance or video-surveillance, region tracking and segmentation for video compression, gestures tracking in humancomputer interaction. In mobile robotics, visual tracking methods are necessary in many tasks, for example in people or objects tracking for obstacles avoidance, and landmarks tracking for autonomous navigation.

To deal with the large number of applications many visual tracking methods have been proposed in literature. However, any of them are enough robust to deal with all possible environments conditions and targets presents in such applications. Michel Devy LAAS - CNRS, 7 avenue Colonel Roche 31077 Toulouse, Cedex 04, France <u>michel@laas.fr</u>

Visual tracking task consists in determine a target configuration over an image sequence, when it is in apparent movement on images. Movement could be a real motion on camera visual field of the target or it could be produced by the movement of camera over a turntable, an arm or a robot, or both.

Must of the visual tracking methods has been conceived or proposed to work only on a small set of conditions and/or for specific environments and targets. For example, in specific tasks like video surveillance or pieces assembling is possible to find the best method to deal with such a task, considering a fixed camera, a controlled environment and limited targets.

In dynamic worlds, mobile robots have to make very different and complex tasks which needs from a visual tracking method. In order to deal with different environments conditions and targets it is indispensable to integrate on board more than one visual tracking method. Moreover, it is necessary to select specific targets and environments conditions for which each one of the tracking methods works.

We are interesting on autonomous mobile robot navigation on human environments. In order to localize the robot on a global map we use planar landmarks like posters or emergency signals. The environments conditions are very different, going from cluttered spaces like a library, to contrasting objects on plain walls.

We seek to characterize targets and environments conditions in order to select automatically the bestadapted tracking method. To characterize them we proposed to measure the image complexity over three specific regions, defined from a recognized target.

In section 2, we describe the four tracking methods used, and in section 3 we defined the regions and complexity measure used to characterized targets and environments. In section 4, we present results of the track-

ing strategy, and finally in section 5 we give our conclusion and future work.

#### 2. Tracking methods

We use four tracking methods: a) active contours b) 1D correlation, c) a set of points tracker and d) a template differences tracker.

In order to evaluate tracking methods and their limitations we could analyze them with the following elements [1]:

- Target representation
- Observation space representation
- Hypothesis generation
- Hypothesis measures

#### 2.1. Active Contours

Active contours or snakes have been proposed by [2], as a minimization technique to segment and track objects. We use a variation of the original method as described in [3], which let to distribute control points over the curve, in function of their curvature (figure 1).

In this method *target representation* is given by object's contour. Generally, a target model is not given, but models could be incorporated as new energy terms. In order to track planar landmarks, we used the defined model described in [4].

*Observation space representation* is given by the external potential field commonly the image intensity gradient.

The *Hypothesis generation* are given specifically by the method used to solve the equation of movement, generally, for example, by variational methods, or as in our case by a dynamic programming algorithm [5].

*Hypothesis measure* is given by finding the minimal energy state for each step.



Figure 1. Tracking a poster with templatebased active contours: a) tracking image, and b) Observation space representation.

#### 2.2. 1D correlation

This method has been used to follow simple lines segments, like straight lines or ellipses and also for complexes structures formed from a ser of lines like in [6]. The main idea behind this method is to avoid the explicit line extraction on each image. Considering that a exists a given model of the wanted structure, this method find in certain directions from a given prediction points which could belong to structure, and then validate them to find structure's new position.

So, *target representation* is given by one or a set of lines. As in the previous method, these lines correspond generally to object's contours.

Commonly the *observation space representation* is given by a contour detector like Sobel, Canny or Deriche.

In our implementation the *hypothesis generation* and *measure* are made by a random process called RANSAC.



Figure 2. 1D correlation method: a) doted line is the initialization model, perpendicular lines are the search directions and points are points with maximal correlation and, b) the proposed hypothesis given and tested by RANSAC algorithm.

#### 2.3. Set of points tracker

This method has been proposed in [7]. The tracking is made using a comparison between two set of points. On one hand points extracted from a given region in present image, like discontinuity or interested points, and on the other hand a set of points from a given model or target found on previous image.

*Target representation* is the set of points of the previous model. This target model could evolve as the image sequence flow, adapting the model to new conditions.

*Observation space representation* is given by the discontinuity or interest point detector, as Canny, Deriche or Harris.

In our implementation *hypothesis generation* is made from a given research zone predefined which it is move in a spiral way, along the last known position.

The *hypothesis measures* are done with the Hausdorff distance.



Figure 3. Set of points tracker: a) Target model, b) Tracked target (inner rectangle) and research zone (exterior rectangle).

### 2.4. Template differences tracker

Many templates differences tracking methods has been proposed over the last years, from which the most performances are the one proposed by Belhumeur and Hager in [8] and the one proposed by Jurie and Dhome in [9]. Basically, these methods use the differences from the position of a given template in image at time twith the image at instant t +1, to correct position of template over the last image. We have used the second approach because it has the advantage to avoid the *jacobian* matrix computation, but however it requires previous template learning.

On this method *target representation* is made by a regular sampling of the given target.

Observation space representation is the template differences space. And hypothesis generation and measure is given by the hyperplane approximation, described in [9].



Figure 4. Templates difference tracker: a) initial template, b) movement of given template, and c) template differences.

As we have described we have four tracking methods with deal with different target representations, different observation space representations and different forms to generate and measure hypothesis. Accordingly to this, each tracker has a set of optimal conditions to give the best results. For example, on one hand the set of points and template differences tracking methods works on textured surfaces and on the other hand active contours and 1D correlation tracking methods works when the objects contours are well defined.

We want to characterise the set of optimal conditions for a given tracking method in function of the target and environment complexity. So in next section a complexity measure will be defined in order to achieve this task.

# 3. Complexity Measure

In literature there are many image complexity measures proposed. They can be divided in global or local measures as described in [10]. In order to characterize targets and environments it is necessary to use local measures. We use the complexity measure called spatial temperature. Mainly, because it could be easily calculated and second because is very near the human concept of complexity. The spatial temperature is defined as the intensity variation density per unit area (figure 5) and defined the following equation:

$$T = \frac{\sum_{i=1}^{n} \nabla I(pxl_i)}{n}$$
(1)

where *T* is the computed temperature, *n* are the number of pixels per unit area and  $\nabla I$  is the intensity gradient on the given pixel.



Figure 5. Spatial temperature or intensity variations density per area unit: a) Original image, b) temperature's image (T = 35).

Considering that, a visual tracking method is a local procedure to search for a solution in a given neighborhood. We need to characterize targets and environments complexity in their neighbor regions. To measure image complexity in specific regions of interest, three zones have been defined as follows:

- a) Contour zone.- It is a small zone around target contour, between 5 to 10 pixels thickness (size depends on targets and/or applications).
- b) Inner zone.- It is region inside and not contained by the contour zone.
- c) Exterior zone.- It is a limited region outside the contour zone between 10 to 20 pixels strength.



Figure 6. Some examples for different environments conditions and the corresponding temperature zones: in (a), (b), (c) and (d) we show the different contexts and the three zones where temperature is calculated, the inner rectangle is the contour zone, inside is the inner zone and between the two rectangles is the exterior zone, in (e), (f), (g) and (h) we show the gradients found for each zone in each context.

We have measure spatial temperature for each one of three zones and consequently we have characterized complexity conditions where each one of the tracking method works and gives good results.

In figure 6, we show different images of targets in different contexts and the three defined zones (see figure text) where spatial temperature is going to be measure. In (a) our target has a white border and as we can see in the corresponding image of the three zones in (e), there are a few gradient points outside de object's contour, so temperature in this zone is near to zero. In other words, this case is very similar to the one in (d) and corresponding temperature zones (h), where our target is on white environment, so complexity of environment in both cases (a) and (d) does not affect tracking methods.

In (b), our target doest have a the white border so complexity of environment is taken into account on exterior temperature as we can see in corresponding image (f), and finally in (c) and (g) are show another one of the targets used. In table 1, it is show the temperature measure T of the three zones for each one of the top images on figure 6. As we can see on table, for target on image (b), the temperature on three zones are very similar, so complexity on this scene and then complexity of the tracking is reflected by the fact that three zones have higher temperature. On target from image (d) the inner and exterior temperature are the lowest, so we hope that tracking methods that use tar-

get representation as object's contour works very well on these conditions.

(a)         (b)         (c)         (d)           Contour         24.93         29.47         18.99         18.43           Inner         20.84         29.44         17.71         12.73	Zones	Targets			
Contour24.9329.4718.9918.43Inner20.8429.4417.7112.73		(a)	<b>(b)</b>	(c)	( <b>d</b> )
<b>Inner</b> 20.84 29.44 17.71 12.73	Contour	24.93	29.47	18.99	18.43
	Inner	20.84	29.44	17.71	12.73
<b>Exterior</b> 1.85 26.22 1.06 2.76	Exterior	1.85	26.22	1.06	2.76

Table 1. Zone temperatures for targets showon figure 6.

# 4. Set of complexity conditions for tracking methods

In order to have the set of conditions where each one of the tracking methods gives good results, we have tested many targets on different conditions of backgrounds. The results are show on table 2. As we can see on this table, the set of optimal temperature conditions for a good tracking with active contours is that the temperature of contour zone need to be higher than other ones, and it is expressed as the ratios between inner and contour and external and contour temperatures. It follows that, when exists a very low temperature on contour zone and near zero temperature on

other zones, active contour will still work. The reason is that active contour are conceive as an energy minimization method, ant it will be attracted still by weak potential fields, in not complex environments. The case of 1D correlation tracking is very similar, but, on this case the gradients presents on observation space need to be stronger, because on RANSAC method there is a threshold to consider that point is a part or not of the hypothesis. This is expressed by the condition  $T_c > 12$ .

Tracking method	Set of temperature conditions	
Active contours	$rac{T_{i}}{T_{c}} < 0.75 \ { m and} \ rac{T_{e}}{T_{c}} < 0.75$	
1D correlation	$\frac{T_i}{T_c} < 0.68, \ \frac{T_e}{T_c} < 0.52 \ \text{and} \ T_c > 12$	
Set of points	$T_i > 10, T_c > 8 \text{ and } rac{T_e}{T_c} < 0.48$	
Template differences	$rac{T_i}{T_c} > 0.54$ , $T_i > 14$ and $T_e \approx cte$	

## Table 2. Set of temperature conditions for each tracker. Subscripts on T represent contour, inner and external temperatures zones.

The set of points tracker, needs high inner and contour zone temperatures, mainly because target model need interest points on both inner and contour zone. Finally, for template difference tracker, it is needed a high inner temperature, because this method use texture inside the target and the exterior zone could have some complexity, but it need to be constant, that because this method use a learning phase with initial conditions that need to be respected along the tracking

# 5. Conclusion and future work

We have presented a methodology to characterize complexity of targets and environments for selection of tracking methods. The complexity characterization is based on measure of spatial temperature over three predefined zones: a) over the targets contour, b) inside the target and c) on a small neighborhood outside the target. A set of optimal temperature conditions has been obtained for each one of the tracking method. This information is used to select the tracking method that better works on the given set of temperature (complexity) conditions. In case of changing conditions, as for example the movement of a target over different background, this information is used to select the optimal tracker for new temperature (complexity) conditions. In future work new targets and tracking methods will be incorporated in order to deal with 3D targets and with changing illumination conditions.

#### References

- Wu Y., And T. S. Huang, "A Co-inference Approach to Robust Visual Tracking", in Proc. of IEEE *Int'l Conf. on Computer Vision* (ICCV'2001), pp. 26-33, Vancouver, Canada, 2001.
- [2] Kass, M., A. Wirkin, and D. Terzopoulous, "Snakes: Active Contours Models", in Int'l J. Computer Vision, vol. 1, pp. 321-331, 1988.
- [3] Marin-Hernandez, A., and H.V. Rios-Figueroa, "Eels: Electric Snakes", in *Computación y Sistemas*, vol. 2, no. 2-3 pp. 87-94, 1999.
- [4] Marin-Hernandez A. and M. Devy; "Model-Based Active Contour for Real Time Tracking" In Proceedings of *International Symposium on Robotics and Automation* 2002; Toluca, Mexico, September 1-4, 2002.
- [5] Amini, A. A., T.E. Weymouth, R.C. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision", in IEEE. *Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, Sep. 1990.
- [6] Drummond, T. and R. Cipolla, "Real-Time Visual Tracking of Complex Structures", in IEEE. *Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, No. 7, pp. 932-946, July, 2002.
- [7] Huttenlocher, D. P. et al, "Visually guided navigation by comparing two dimensional edge images", *Technical. Report*, TR-94-1407, Stanford University, Stanford, California, January, 1994.
- [8] Hager, G. D. and P. N. Belhumeur, "Efficient Region Tracking With Parametric Models of Geometry and Illumination", in IEEE. *Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, No. 10, pp. 1025-1039, October 1998.
- [9] Jurie, F. and M. Dhome, "Hyperplane Approximation for Template Matching", in IEEE. *Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, No. 7, pp. 996-1000, July 2002.
- [10] Peters, R. A. II, and R. N. Strickland, "Image Complexity Metrics for Automatic Target Recognizers", in Proceedings of Automatic Target Recognizer System and Technology Conference, Silver Spring, USA, October, 1990.