A Probabilistic Approach for Adaptive Sizing in Visual Tracking

Héctor Barrón, Janeth Cruz, Leopoldo Altamirano

National Institute of Astrophysics Optics and Electronics
Luis Enrique Erro No. 1 Tonantzintla, Puebla, México
{hbarron,jcruzæ,robles}@inaoep.mx

Resumen. En este documento se presenta un método de seguimiento adaptable, el cual es robusto al basarse en métodos probabilísticos para estimar el estado correcto del objeto en seguimiento. La mayoría de los trabajos relacionados solo consideran la dinámica del objeto; sin embargo, el método propuesto es adaptable también a características espaciales tales como el tamaño del objeto. Para estimar la posición y el tamaño del objeto se utiliza un modelo probabilístico markoviano tal como el Filtro Kalman. El método mantiene el seguimiento aún cuando el objeto de interés se encuentra afectado por una oclusión o fondo ruidoso. El desempeño del seguimiento es probado con secuencias de imágenes sintéticas y reales, donde se mide la precisión y la exactitud bajo diferentes condiciones.

Palabras Clave: Seguimiento visual, tamaño adaptable, MAP

Abstract. In this paper we present an adaptive tracking approach that is strengthened with probabilistic methods to estimate the correct state of the object. Most of the tracking approaches are focused only on dynamic (temporal) behaviors of the target, however our method is also adaptive to spatial characteristics of the target such as its size. Because of that, the size and the shape of the object can change over time. To estimate the target position we use a Probabilistic Markov Model as Kalman Filter, and to estimate the object appearance we use the same method defining a search space with random samples. Furthermore, the method maintains the tracking even when the target is under occlusion or noisy background. Tracking performance is tested with synthetic and real image sequences and we present the precision and accuracy under different conditions of tracking.

Keywords: Visual Tracking, Adaptive Sizing, MAP

1 Introduction

Movement-based analysis from images has a wide range of application in the medicine, industry and military context. Tracking is one of the strategies more used in Computer Vision for information extraction from objects of interest. A tracking system collects a set of measurements from sensors, these measurements represent target attributes. Nevertheless, visual tracking is one of the issues more difficult because it is affected by different conditions present in dynamic environments as such changes in object appearance, sensor or environment noise and clutter background that affect the measurements quality. This situation is intensified when partial or total occlusion appears at scene.

Tracking can be seen as an image registration problem, where the information about the object behavior can be useful to obtain robust performance. We proposed an adaptive tracking method based on probabilistic approach where the information obtained from measurement determines the search space. Another contribution in this work is the implementation of an adaptive update of the reference image.

We start by reviewing the problem of visual tracking as an image registration problem and some of the most relevant methods in section 2. The proposed method is described in section 3. Section 4 presents the results on synthetic and real image sequences. Finally, some conclusions are given in section 5.

2 Previous Work and Context

Tracking can be seen as image registration problem where it is necessary to find a mapping between two images of the same object at different times, limited by a tracking window. The image of the object at the previous time is the reference image and the image of the object at the current time is the input image. It is necessary to identify three main components for solving the image registration problem [1]:

- A search space defined by a set of features to measure at each image.
- A similitude measure between two set of features.
- A search strategy for the set of values that optimize the similitude measure.

The features to consider for measuring can be some control points or pixel values, but this depends of the kind of object to track and the scene conditions. We can find several papers where tracking is based on template schemes because the kind of target has a textured layout given for the pixel intensities [2].

Several object tracking algorithms have high probability of missed target because of dynamic environments (e.g. illumination changes) or varying-time state of the target in the scene (e.g. position and size). Some authors solve the previous problem using an adaptive tracking window to maintain the tracking. For the reason that, if the tracking window is smaller than the target causes the aperture problem, but if the window is larger than the target, it is possible to have erroneous information about the target. Another problem in tracking is total or partial occlusions since they cause the false alarms in measurement process.

Chien et. al. [7] proposed an adaptive tracking window using a set of conditions to determine the values in a sizing vector. This rules are based on the difference of mean of pixel values and different sections in the tracking window. In Son et. al. work [8], the tracking window is adapted using information about gradient between neighbor pixels. This method is more flexible than in [7] but it is necessary that the target displacements were small as 1-2 pixels per frame.

Another approach consists in increasing the search space as much as freedom degrees could have the object. Collins [9] proposed to perform a search in locations space and another in scale space. His method uses the histogram as set of features. This scheme can track an object with changes of size and displacements larger than 2 pixels per frame, but the object behavior information is not used.

It is possible to review another papers [4,5] where probabilistic method are proposed to exploit the information obtained from the object behavior. In [5], it is proposed to use a set of random samples around a prediction on geometrical parameters of the object and to use a probabilistic density as similitude measure. This algorithm needs an extension to maintain the tracking during occlusion. Ross et. al. [4] proposed a similar method than [5] based on the maximum a posteriori estimate but it is used an adaptive reference image based on eigenbasis to add new visual features to the object. The work described in this paper is based on the latest approaches.

3 Proposed Adaptive Tracking Method

In this section, we describe the proposed tracking algorithm and theoretical basis is discussed under context.

3.1 Probabilistic Model for Tracking

Tracking can be seen as the estimation of the state of a moving object based on measurements [3]. In visual tracking, those measurements need to be taken from images, so let X_t be the object state that can be determined by the sequence of images $I_t, I_{t-1}, I_{t-2},...I_0$.

Meanwhile, due the object behavior across the time, it is possible to estimate the object state at an specific point in time. Let X_t be the object state in the time t, we can consider a pdf (probability density function) to model the object state, such as let $p(X_t|X_{t-1})$ be the probability of the object state where X_t is just determined for X_{t-1}. So, we can establish the priori knowledge about the object behavior.

A method to obtain an estimate of the object state that allows incorporate new measurements from images with the prior knowledge, is given for Bayes' rule. Given the current image I_t and the previous state X_{t-1}, X_t can be found using the Maximum a Posteriori estimate, i.e.,

$$X^*_t=\mathrm{argmax}_X\, p(X_t|I_t,X_{t-1})\,, \tag{1}$$

$$p(X_t|I_t,X_{t-1}) \propto p(I_t|X_t)p(X_t|X_{t-1})\,. \tag{2}$$

The measurement process is a search for the maximum likelihood $p(I_t|X_t)$ in the neighborhood of X^*_t and the set of possible states are derived as a measurement pixels. This method is more flexible than in [7] but it is necessary that the target displacements were small as 1-2 pixels per frame.

Another approach consists in increasing the search space as much as freedom degrees could have the object. Collins [9] proposed to perform a search in locations space and another in scale space. His method uses the histogram as set of features. This scheme can track an object with changes of size and displacements larger than 2 pixels per frame, but the object behavior information is not used.

It is possible to review another papers [4,5] where probabilistic method are proposed to exploit the information obtained from the object behavior. In [5], it is proposed to use a set of random samples around a prediction on geometrical parameters of the object and to use a probabilistic density as similitude measure. This algorithm needs an extension to maintain the tracking during occlusion. Ross et. al. [4] proposed a similar method than [5] based on the maximum a posteriori estimate but it is used an adaptive reference image based on eigenbasis to add new visual features to the object. The work described in this paper is based on the latest approaches.

3 Proposed Adaptive Tracking Method

In this section, we describe the proposed tracking algorithm and theoretical basis is discussed under context.

3.1 Probabilistic Model for Tracking

Tracking can be seen as the estimation of the state of a moving object based on measurements [3]. In visual tracking, those measurements need to be taken from images, so let X_t be the object state that can be determined by the sequence of images $I_t, I_{t-1}, I_{t-2},\ldots I_0$.

Meanwhile, due the object behavior across the time, it is possible to estimate the object state at an specific point in time. Let X_t be the object state in the time t, we can consider a pdf (probability density function) to model the object state, such as let $p(X_t|X_{t-1})$ be the probability of the object state where X_t is just determined for X_{t-1}. So, we can establish the priori knowledge about the object behavior.

A method to obtain an estimate of the object state that allows incorporate new measurements from images with the prior knowledge, is given for Bayes' rule. Given the current image I_t and the previous state X_{t-1}, X_t can be found using the Maximum a Posteriori estimate, i.e.,

$$X^*_t=\mathrm{argmax}_X\ p(X_t|I_t,X_{t-1})\ , \tag{1}$$

$$p(X_t|I_t,X_{t-1}) \propto p(I_t|X_t)p(X_t|X_{t-1})\ . \tag{2}$$

The measurement process is a search for the maximum likelihood $p(I_t|X_t)$ in the neighborhood of X^*_t and the set of possible states are derived as a measurement vector Z. So $p(I_t|X_t)$ describes the probability of having an image with the attributes defined by X_t and dependent of the measurement parameters, under the probabilistic distribution given for $p(X_t|X_{t-1})$. We can define the search space as the whole set of possible values for the measurement parameters in vector Z. They can be geometric attributes, lighting conditions, camera parameters, etc. These variables are related to a state vector X with a measurement function H(X)=Z.

We can solve this recurrent process, given p(X_t|X_{t-1}) is Gaussian, with a Probabilistic Markov model as Kalman Filter. Our proposal is established by a state vector X_t defined by $[x, y, s, v_x, v_y, v_s]$, where (x,y) is the location of the object, s is the scale of the object and $[v_x, v_y, v_s]$ is the change rate between one frame and the next. The measurement vector Z is given by [x, y, s] and $p(I_t|X_t)$ is approximated by

$$p(I_t|X_t) = \mathrm{sig}\left(\frac{\sum_{x,y} T(x,y)I(x,y) - M\bar{I}\bar{T}}{\left(\left(\sum_{x,y} T^2(x,y) - M\bar{T}^2\right)\left(\sum_{x,y} I^2(x,y) - M\bar{I}^2\right)\right)^{1/2}}\right), \tag{3}$$

where *sig* is the sigmoid function, T(x,y) represents the reference image, I(x,y) represents the input image, M is the number of pixels in each the image, T and I are the mean of the reference and input image, respectively.

3.2 Adaptive Reference Image Scheme

The estimate of the object state is based on a recursive Bayesian method to deal with time-varying data, so the update of the reference image must be based on a similar adaptive method to make robust in front of false registrations and drift problems. We proposed an approach where the reference image gradually changes as much as the object appearance changes in the image.

Each pixel is modeled as a Gaussian density where the whole set of mean values represents the current state t of the object surface:

$$p(X,\mu_t,\sigma_t) = \frac{e^{(X-\mu_t)^2/\sigma^2}}{\sqrt{2\pi\sigma^2}}, \tag{4}$$

where μ_t is the mean and σ^2 is the variance of the density. At each time, the state of the object surface is estimated, due the pixel value as a measurement. If we used a set of Gaussian we would model a multimodal appearance of the object [6], but not in this paper. The on-line update of the parameter μ_t is given by:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t, \tag{5}$$

$$\rho = \alpha \, p(X,\mu_t,\sigma_t) \, , \tag{6}$$

α is defined as a time-varying gain or an adaptive constant. $p(X,\mu_t,\sigma_t)$ can be approximated to 1 if X_t is near to mean μ_{t-1}, otherwise $p(X,\mu_t,\sigma_t)$ is 0.

3.3 Tracking Algorithm

An overall block diagram of our proposed method is shown in Fig 1. Due the concepts reviewed at the section 2, two image registration problems need to be solved. First, we measure the localization of the object with a search space defined by a prior estimation and a search strategy considering every possible position of the object. Second, we measure the scale of the object, due an scale estimation and a set of random sampled defined by a Gaussian density.

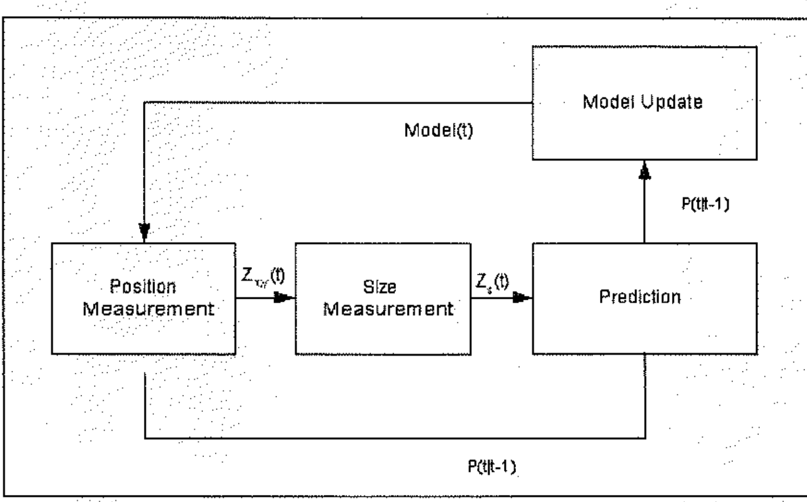


Fig. 1. General diagram of the adaptive tracking algorithm

The resulted measurements are used to estimate the object state and the reference image is adapted to the visual conditions at the current frame, due the prediction in previous time. The adaptive tracking algorithm consists in the following points:

1. Given an image sequence, the target is selected manually where the first reference image and the state of the object are defined. Alternatively, we can use a segmentation algorithm to select automatically a moving object.
2. In the Prediction module, an estimate of the position and the scale of the object are calculated. So, we define our search space for the next frame, around this estimate.
3. At the next frame, the object position is measured by the Position Measurement module using the similitude measure defined in section 3.1. For this case, the

object position with the maximum a posteriori probability is selected. We use an sequential search strategy.
4. Due the position measurement as starting point, the scale measurement is obtained in the Size Measurement module. Using a set of random samples we reduce the search space for selecting the scale with the maximum a posteriori probability.
5. Using the pixel values of the image defined by the prediction at time t-1, the image of the object model is adapted using the method defined in section 3.2.
6. Finally, go to step 2.

4 Experimental Results

To evaluate the performance of the proposed algorithm for the adaptive sizing of a tracking window, in scenes with target under partial and total occlusion, we use synthetic and real image sequences. Three different basis synthetic sequences with 120-180 frames were created, where a target maintains linear or nearly linear movements with displacements between 2-8 pixels. The target size changes from 15×12 to 110×90 pixels and the change rate of size is from 1 to 4 pixels per frame. the target presents partial and total occlusion.

We applied Gaussian noise to each sequence with five different signal-to-noise ratios (SNR) defined as follows:

$$\mathrm{SNR} = 20\log\left(\frac{|\mu_t - \mu_b|}{\sigma_n}\right) dB \ , \tag{7}$$

where $|\mu_t\text{-}\mu_b|$ is the absolute difference of the intensity average between the target and the background and σ_n is the standard deviation of the added Gaussian noise. To evaluate the performance, we used the size error measurement defined by:

$$\mathrm{E}_s = \frac{|A - A_R|}{A_R} \ , \tag{8}$$

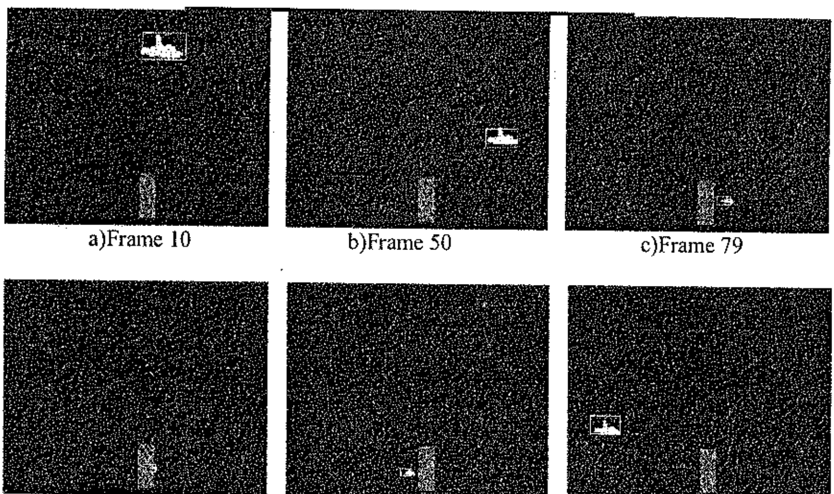
where A is the area of the tracking window measured by the system and A_R is the real area of the target.

In Fig. 2, performance of the proposed tracking algorithm over a synthetic sequence is shown, where the SNR value is 6dB. The target is affected by size changes along of the sequence, as well as occlusions (Fig. 2.d). To maintain the tracking even under occlusion, it is necessary to have an estimate of the object appearance when the object appear again, as such our method does. In Fig 3 is shown the tracking performance over change of size.

In Table 1, statistics of size error (E_s) on synthetic sequences are shown with varying SNR from 0.0 to 10.0 dB. The average error μ_E describe accuracy and standard deviation σ_E shows precision of our adaptive algorithm.

Table 1. Stadistics of size error. This table shows how accuracy of the estimate is decreased as noise is increased

SNR	Mean Error	Stand. Dev.
10	0.2821	0.0567
8	0.3948	0.0785
6	0.3057	0.0795
4	0.4798	0.0975
2	0.4232	0.0921



a)Frame 10 b)Frame 50 c)Frame 79

d)Frame 84 e) Frame 91 f) Frame 123

Fig. 2. Sampled tracking example using the proposed method for a synthetic sequence with SNR=6dB

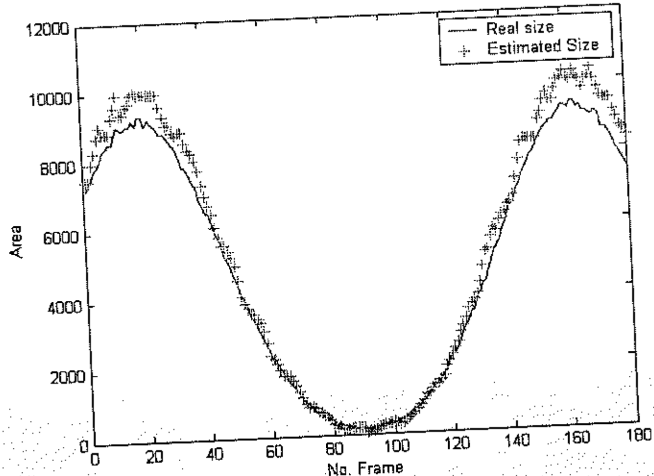


Fig. 3. Adaptive sizing in tracking. Here is illustrated how the size estimate is made correctly, allowing the size changing of the object model (Reference image)

To evaluate the proposed algorithm with real environments, six different sequences were used. Different objects were selected as target where occlusion and size changes appear. Objects have linear and nearly linear movements with some atmospheric factors, such as snow and haze. In Fig. 4 is shown tracking over sequence *Winter*. The frame 40 (Fig. 4.b) presents an object under total occlusion, also the target size changes along the sequence, which increases the difficult to maintain the tracking, but our algorithm resolve this trouble.

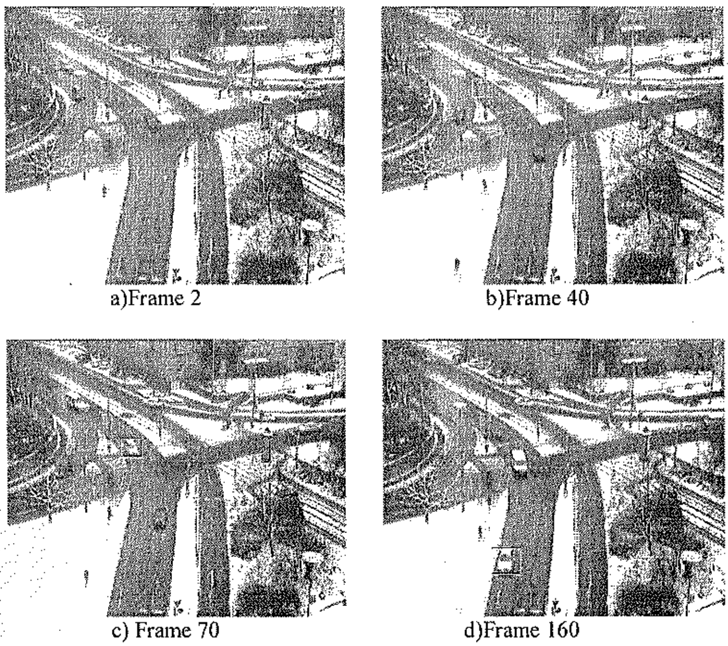


Fig. 4. Real image sequence. Example of tracking when using proposed algorithm with real sequence Winter. Snow and occlusion appear in scene

We demonstrated the reliability of our method in Fig. 5 where it is compared with two schemes more. A dynamic strategy consists in changing the reference image at each frame and a static scheme consists in maintain the reference image without changes along the sequence. To evaluate target position error, we use the absolute difference between real position and the estimated position in our algorithm.

Tracking with dynamic strategy over this sequence is affected by the drift problem induced by the discrete nature of the visual tracking. Meantime, when the object changes its appearance, the static strategy is obsolete. The adaptive strategy maintenance the tracking to the end of the sequence.

In Fig 6, the performance of the three techniques is shown over a real sequence where haze appears. The object changes its appearance in the sequence. The static strategy quickly missed the target, but the dynamic strategy seem not to miss the target.

In the Fig. 7 we observe the visual results over this sequence. The Fig. 7.a shows the initial position and initial size of the target. Fig 7.c shows as the dynamic strategy lost the original object and it actually follows a corner of the object. However, the adaptive strategy maintains the tracking over the original object.

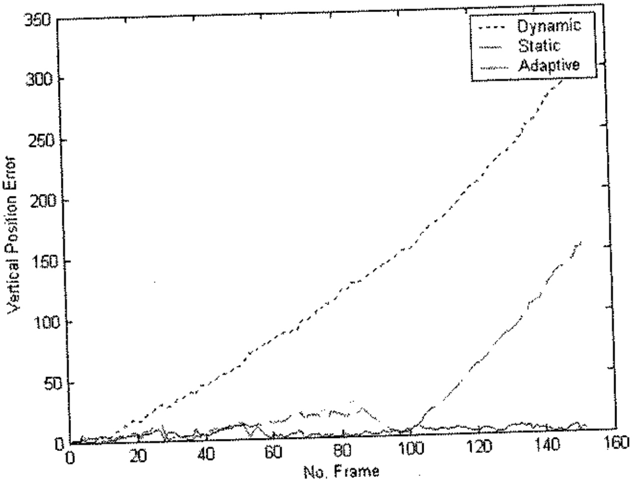


Fig. 5. Comparisons between dynamic and static schemes and our adaptive algorithm on sequence Winter

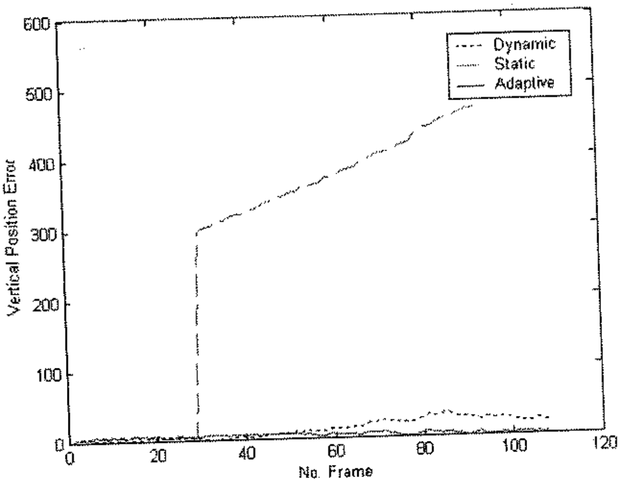
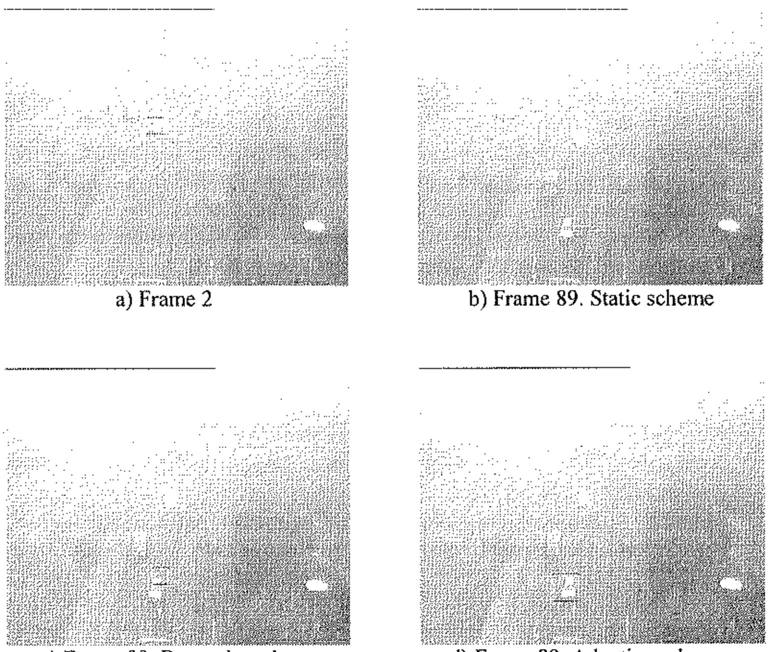


Fig. 6. Comparisons between dynamic and static schemes and our adaptive algorithm on sequence Nebel



a) Frame 2

b) Frame 89. Static scheme

c) Frame 89. Dynamic scheme

d) Frame 89. Adaptive scheme

Fig. 7. Visual results on sequence Nebel, using b)static scheme, c) dynamic scheme and d) Adaptive scheme

5 Summary and Conclusions

We have presented a visual tracking method based on probabilistic methods where tracking is seen as an image registration problem, where the matching between the reference image and the input image is defined by the estimate over the target behavior. Results of experiments over synthetic and real video sequences ensure that our proposal is reliable when the object size changes and occlusions are presented.
We presented comparisons with two techniques more to update the reference image and the adaptive method obtained satisfactory results. Our work was tested over sequences with dynamic environments and Gaussian noise. The tracking window is adapted to size changes but not to rotation movements. It is possible to extend this work to multiple targets with non-lineal motion using different estimation filters (e.g. Particle Filter).

References

1. Brown, L. G.: A survey of image registration techniques. ACM Computing Surveys. Vol. 24, No. 4. (1992) 325-376.
2. Eklund, M. W., Ravichandran, G., Trivedi, M. N., Marapane, S. B.: Real-time visual tracking using correlation techniques. Proceedings of the Second IEEE Workshop on Applications of Computer Vision. (1994) 256-263.
3. Bar-Shalom, Y., Rong Li, X., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation. John Wiley and Sons. 2001.
4. Ross, D., Lim, J., Yang, M. H.: Adaptive Probabilistic Visual Tracking with Incremental Subspace Update. ECCV 2004, Vol. 2. (2004) 470-482.
5. Rasmussen, C., Hager, G. D.: Probabilistic Data Association Methods for Tracking Complex Visual Objects. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 6. (2001) 560-576.
6. Power, P. W., Schoonees, J. A.: Understanding Background Mixture Models for Foreground Segmentation. Proceedings Image and Vision Computing New Zealand (2002). 267-271.
7. Chien, S. I., Sung, S. H.: Adaptive window method with sizing vectors for reliable correlation-based target tracking. Pattern Recognition, Vol. 33 (2000) 237-248.
8. Son, J. G., Lim, C. W., Choi, I., Kim, N. C.: Adaptive Sizing of Tracking Window for Correlation-Based Video Tracking. IEICE Trans. Inf. and Syst. Vol. E85-D (2002) 1015-1021.
9. Collins, R. T.: Mean-shift Blob Tracking through Scale Space. Computer Vision and Pattern Recognition (CVPR'03), IEEE, June. (2003).