

# About Some Methods of Motion Detection and Estimation: A Review

Amara Kahina, Djekoune A. Oualid, Belhocine Mahmoud, and Achour Nouara

**Abstract**—Video surveillance systems have known significant growth because of the increased insecurity in these recent years. In order to reduce threats such as assaults, many cameras have invaded the public squares. The manual monitoring of these screens is tedious because of the large amount of information. So it is very interesting to automate this process from image processing systems able to extract the useful information from video sequences and interpret it. One of the most important tasks is the motion detection and estimation. This article aims to provide the status of art of the different techniques of motion detection estimation and segmentation based on movement. Many studies have been conducted on the subject and the literature is very abundant in this province, we are not trying to list all the existing methods. The idea is to give an overview of the most commonly used methods and to distinguish different types and approaches.

**Index Terms**—Motion detection, motion estimation, segmentation based on movement, video surveillance.

## I. INTRODUCTION

The unusual motion detection is the first step in many video surveillance applications. Many Applications in this province have been developed to automate the monitoring task completely or partially with high sensibility and a false alarm rate as less as possible.

Used in many situations, usually for security reasons:

- In road safety subject, using specialized cameras or sensors to assess the density of traffic slowdowns that may arise, the presence of people on the emergency stops bands [1], [2].
- For monitoring the sensitive locations (banks, nuclear power plants, etc.), to prevent trespassing, burglary (theft, assault) [3], [4].

The motion detection is the identification of pixels in each image belonging to moving objects. The motion estimation is a quantification of the simple movement, by translation vectors, pixel-oriented, objet or block.

The Segmentation based on movement is a task that goes beyond detection as it is to segment each image into regions that represent homogeneity of apparent motion (Fig. 1). These approaches are used in various fields such as video compression, motion tracking, semantic analysis of the scene, etc.

Manuscript received December 5, 2013; revised March 31, 2014.

Amara Kahina is with CDTA, Advanced Technologies Development centre Baba Hassen.BP 17, Baba Hassen, Algiers, Algeria (e-mail: amara.kahina1988@yahoo.com).

Djekoune A. Oualid, Belhocine Mahmoud, and Achour Nouara are with Institute of Electronics and Computer Science, USTHB, University of Science and Technology Houari Boumediene BP32, USTHB, Bab Ezzouar, Algiers (e-mail: odjekoune@cdta.dz, mbelhocine@cdta.dz, nouara.acchour@gmail.com).

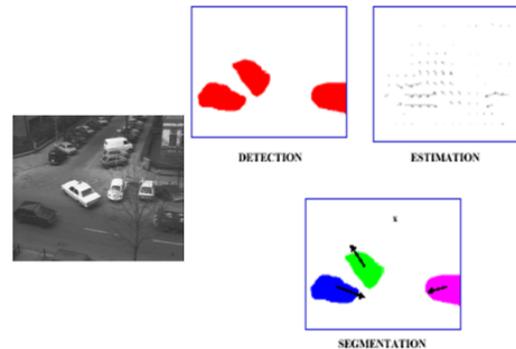


Fig. 1. Detection, estimation and motion segmentation [5].

## II. MOTION DETECTION ESTIMATION

Motion detection is a preliminary task to many treatments of the highest level in computer vision such as compression or tracking [6]. The computer does not have a semantic interpretation of the scene. This interpretation is however an essential ingredient in the process of detection of moving objects. Detect motion in a sequence of images is not a new problem, but still open today.

The set of the seen status of art allows us to establish a hierarchical classification of different methods of motion detection.

This classification is based on the background modeling, so we identify four main methods:

### A. Detection without Modeling the Background

The principle of these methods is to detect motion by calculating a mathematical quantity which is a function of the intensity or a color at any point of the image. This amount is supposed to reflect the importance of visible movement in the scene. The time derivative of the light intensity [7], the spatiotemporal entropy of the image [8] and the standard optical flow are included in this category.

The Fig. 2 shows the result of motion detection by the time derivative.



Fig. 2. Detection by the difference between consecutive images [9].

### B. The Local Modeling of the Background

These methods involve at any point of the image a value or function to model the appearance of the background at this point. The appearance model of the background at a point depends only on the observations occurred in this place at this

point, the other image pixels are not involved. The simplest model of the background would be a picture which shows a scene devoid from objects [10]. Applying detection by modeling the background with an image leads to results shown in Fig. 3.



Fig. 3. Detection by local modeling of the background [11].

The vast majority of methods presented in the literature build a statistical model, but it may be a stochastic process, a predictive filter or just an intensity value.

### C. Semi-Local Modeling of the Background

These methods are very similar to the previous category; with the close difference that the modeling of the background at a point depends on the observations that took place in a neighborhood of this point, or region of the image to which it belongs. The semi-local modeling includes the detection region [12], the characterization by texture [13] and the posteriori regularization.

### D. Global Modeling of the Background

These methods use every moment all the observations to construct a model of the background. Modeling by switching between several models [14] and a vector space modeling [15] are among the approaches in this category.

The motion detection methods based on the intensity of pixels have limitations to the changes of illumination. As the time derivative, this method shows a little robust to the phenomena of slow or jerky movements, stops short of moving objects, or the presence of some redundant frames in video sequences. It is therefore necessary to perform a temporal smoothing of the sequence that means to apply a moving average operator to the obtained measurement.

While the methods are seeking to model the background having the advantage of detecting both slow movements that fast movements. Moreover, even momentarily stationary objects are detected. However, external environments, the variations in light intensity make rapidly obsolete such model.

Otherwise, it is not always possible to obtain a completely empty image of the scene, under these conditions, it is necessary to update the background image.

## III. MOTION ESTIMATION

The motion estimation is used in image processing to identify moving objects and further define their position and their speed. In video compression, it is part of the prediction process that tries to exploit the temporal redundancy of previously encoded images to predict the movement and thus compress data more efficiently than by spatial prediction. In robotics, this method allows to predict the shifting or position of objects in order to ameliorate the interaction model with the external environment.

We mention methods that return primarily in the motion estimation:

### A. The Block Matching

The block-matching (BM) is a method which allows to estimate the motion in an image stream, it consists in partitioning the image into blocks of equal size, then to search how do the blocks move between two successive frames  $n$  and  $n+1$  (Fig. 4) and then to find the most similar block to the initial block. In this approach, we seek the "best match" between two spatial windows (or blocks). This correspondence can be done by minimizing the square error between the illuminations corresponding to respective pixels of the two blocks. So we deduce the corresponding displacement of the initial block.

All pixels of the same block will have the same movement; this is due to the assumption of homogeneity of the movement on a small enough areas.

The block matching is the most essential form of motion estimation by a field of translation vectors between two frames and the most basic but also one of the oldest and most effective methods to estimate the motion. However it does not use any contextual information in the image and has no signification in this sense.

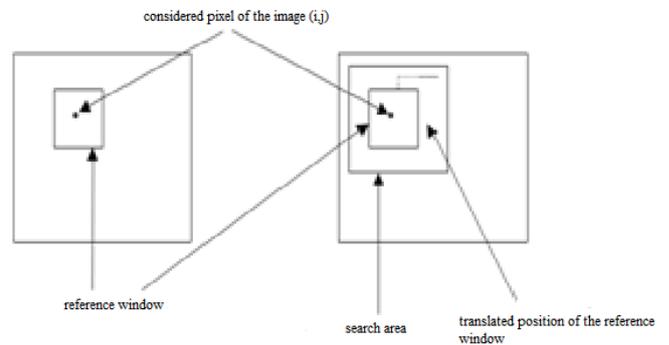


Fig. 4. The Block matching base on a neighborhood of the block size.

### B. The Optical Flow

The goal of the optical flow is to estimate the motion of each pixel of the image within a sequence, as shown in Fig. 5.

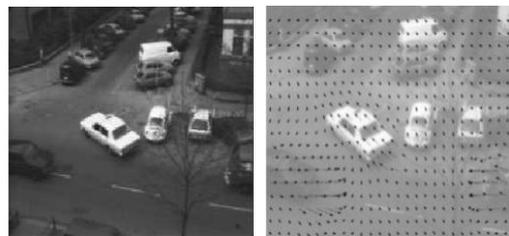


Fig. 5. Velocity field (taxi) [16].

Each image point is determined, at the accurate time by its "color function" (luminance + chrominance)  $I_i(x(t))$ .

The value of the optical flow at time  $t$  for an image point  $x(t)$ , is defined as the movement speed of the image point:

$$v = (v_1, v_2) = \left( \frac{dx_1}{dt}, \frac{dx_2}{dt} \right) \quad (1)$$

The optical flow determination is reached basing on assumptions which are to consider that the illumination is

constant over the image in the analyzed sequence, it results:

$$I(t, x(t)) = I_0 = cte \quad (2)$$

The derived particle function Color  $I_t(x_1, x_2)$  along the optical flow  $(v_1, v_2)$  gives:

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial x_1} \cdot v_1 + \frac{\partial I}{\partial x_2} \cdot v_2 = 0 \quad (3)$$

Assuming that the displacement has low amplitude, the intensity conservation of the assumption linearization leads to the equation of the apparent flow constraint. It therefore has only a single scalar constraint to solve (3) with two unknowns  $(v_1, v_2)$ : the problem of the opening.

The way to find a single solution that solves this equation is to introduce an additional assumption; this method is common to all the optical flow resolution.

At this point, several methods have been exploited:

- Mapping methods;
- Frequency methods;
- Differential Methods;
- Rapid resolution of optical flow with wavelets.

### 1) Mapping methods

The meaning of the mapping methods is to find the displacement field  $\vec{d} = (dx_1, dx_2)$  that matches the best parts of the scene between two consecutive moments. The mapping is usually done from the sum minimization of the squared difference between the two images at times  $t$  and  $t + 1$ , defined as follows:

$$\sum_{(i,j) \in V_{x_1, x_2}} [I((i, j), t) - I((i, j) + \vec{d}, t + 1)]^2 \quad (4)$$

with  $V_{x_1, x_2}$  is the centered neighborhood around the point  $(x_1, x_2)$ .

These methods are often costly in time calculation. To lower the computational complexity, some authors propose to reduce the search space [17], [18]. These methods are robust and simple to enforce. They apply in most video compression standards. However, due to the discretization of the estimated displacement, they become very inaccurate.

### 2) Frequency methods

These methods inspired by the work of Adelson and Bergen [19] study the equation of the apparent flow constraint in the frequency province.

Considering an image sequence  $I(x_1, x_2, t)$  having the Fourier transform  $\hat{I}(f_{x_1}, f_{x_2}, f_t)$ . In case where objects of the sequence undergo the movement of the translation uniform:

$$(x_1(t), x_2(t)) = (x_{10} + v_1 t, x_{20} + v_2 t) \quad (5)$$

According to the assumption of the optical flow, we have the following relation:

$$I(x_1, x_2, t) = I(x_1 - v_1 t, x_2 - v_2 t, 0) \quad (6)$$

$$\hat{I}(f_{x_1}, f_{x_2}, f_t) = \hat{I}_0(f_{x_1}, f_{x_2}) \delta(f_t + v_1 f_{x_1} + v_2 f_{x_2}) \quad (7)$$

with  $\delta(\cdot)$  is the Dirac measure at 0.

The motion estimation is extracted from (7) by

spatio-temporal filtering [16]. The problem of the opening is circumvented by assuming the constant stream of the support for these filters. This assumption has the effect of smoothing the flow in both space and time.

These methods are divided into two branches depending on whether the estimate is based on energy or the phase of the filtered signal. They give very good results; however they are very expensive in calculation time.

### 3) Differential methods

The Differential methods compute the optical flow from the equation 3. There are two different approaches to solve this equation: the dense approaches that compute the optical flow on the entire image and which called local approaches, those local approaches estimate the motion at each point independently [20]. The problem of the opening is circumvented by adding a spatial regularity assumption in the case of dense methods and consistency of a constraint on a neighborhood in the local methods case.

Horn and Schunk suggested the fundamental principles of optical flow estimation in 1981 [21]. These methods are based on regularization on the whole image.

They define a functional composed of two terms and they seek the vector field  $\vec{v}$  minimizing the functional:

$$E(\vec{v}) = H_1(\vec{v}) + \alpha^2 H_2(\vec{v}) \quad (8)$$

$$H_1 = \iint (\vec{\nabla} I \cdot \vec{v} + \frac{\partial I}{\partial t})^2 dx_1 dx_2 \quad (9)$$

$$H_2 = \iint \|\vec{\nabla} v\|^2 dx_1 dx_2 \quad (10)$$

The term  $H_1$  relies on the assumption of brightness, the term  $H_2$  said regularization term and  $\alpha$  is a fixed parameter empirically determined to weight the effects of term.

Since several authors have defined other terms of regularity.

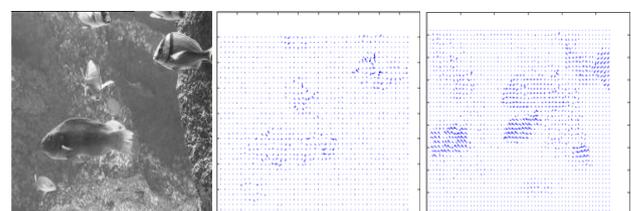
The solution proposed by Lucas and Kanade [22] is to consider that the flow  $\vec{v}$  is locally constant on a neighborhood of the point  $(x_1, x_2)$ . They seek the best flow checking the best the equation of the apparent flow constraint and minimize the following functional:

$$\sum_{(i,j) \in \varphi} W^2(i, j) \left[ \vec{\nabla} I(i, j) \cdot \vec{v}(i, j) + \frac{\partial I}{\partial t}(i, j) \right]^2 \quad (11)$$

with  $\varphi$  is the neighborhood of the point  $(x_1, x_2)$  and  $W(x_1, x_2)$  is a Gaussian.

These methods are based on the regularization of a neighborhood.

Fig. 6 shows the application of the solution proposed by Lucas and Kanade and the motion detection by block matching.



(a) original image (b) LK resulting image (c) BM resulting image  
Fig. 6. Detection by Lucas and Kanade [5].

### C. The Spatiotemporal Filtering

The movement in an image sequence is characterized by a spatiotemporal variation of the light intensity [16]. The spatiotemporal analysis considers a sequence of spatiotemporal data ( $2D + t$ ) a  $3D$  volume which called the spatiotemporal space. The spatiotemporal filtering is based on oriented spatiotemporal filters (often Gabor filters  $3D$ ) for extracting the motion information.

These methods are based on the Fourier transform of moving images. They assume that the optical flow is constant over the spatiotemporal support of the filters, which results a smoothed result in both space and time,

Otherwise, these methods have generally a high computational cost associated with operations of spatiotemporal filtering.

### D. Rapid Resolution of Optical Flow with Wavelets

One of the proposed solutions for solving the optical flow equation is to project it with a wavelet basis on the assumption that the flow is constant [23].

Considering a wavelet basis  $(\varphi^n)_{n=1\dots N}$  and the associated discrete family of wavelets  $(\varphi_{jk}^n)_{n=1\dots N, j \in Z, k \in Z^2}$  defined by:

$$\varphi_{jk}^n(x_1, x_2) = 2^{-j} \varphi^n(2^{-j}x_1 - k_1, 2^{-j}x_2 - k_2) \quad (12)$$

with  $j$ : the index of resolution and  $k$  the index of translation.

The problem of the temporal aliasing encourages us to use the wavelets with large support (large  $j$ ) to estimate large displacements in the scene and wavelets with small support (small  $j$ ) for small velocities.

At a coarse scale  $J$ , we obtain at each point  $(2^j k_1, 2^j k_2)$  the system

$$M_{jk} \cdot V_{jk} = P_{jk} \quad (13)$$

$$\text{with } M_{jk} = \begin{pmatrix} \langle I, \frac{\partial \varphi_{jk}^1}{\partial x_1} \rangle & \langle I, \frac{\partial \varphi_{jk}^1}{\partial x_2} \rangle \\ \vdots & \vdots \\ \langle I, \frac{\partial \varphi_{jk}^N}{\partial x_1} \rangle & \langle I, \frac{\partial \varphi_{jk}^N}{\partial x_2} \rangle \end{pmatrix} \quad (14)$$

$$P_{jk} = \begin{pmatrix} \langle I(t+1) - I(t), \varphi_{jk}^1 \rangle \\ \vdots \\ \langle I(t+1) - I(t), \varphi_{jk}^N \rangle \end{pmatrix} \quad (15)$$

$$\text{and } V_{jk} = (v_1(2^j k_1, 2^j k_2), v_2(2^j k_1, 2^j k_2))^T \quad (16)$$

Solving this problem gives a rough approximation of the movement on a grid of resolution  $2^j$ . The wavelet allows an image structures decomposition according to scaled and spatial location.

## IV. SEGMENTATION BASED ON MOVEMENT

The extraction of motion layers has many applications such as video compression, manipulation of video objects, mosaics generation, semantic analysis of scenes, etc. We separate these methods for estimating motion in different categories, depending on the number of images used simultaneously and according to the temporal distances

separating these images [24].

We distinguish methods working on a pair of consecutive images, those working on a group of consecutive images, and those working on a group of remote images. This classification is summarized in the Fig. 7.

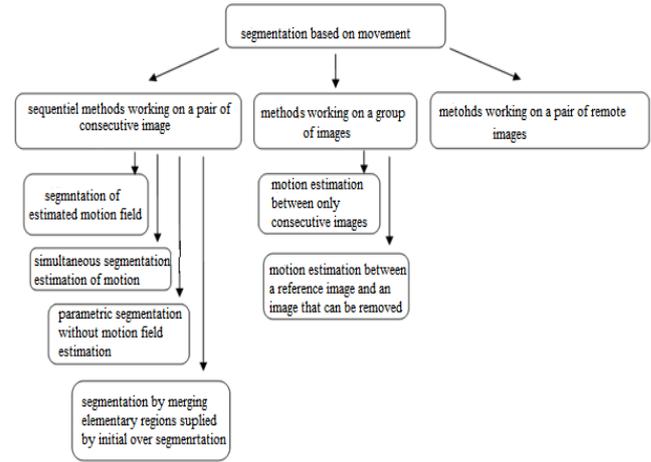


Fig. 7. Classification of segmentation methods based on movement.

### A. Sequential Methods Working on a Group of Consecutive Images

This set contains methods include approaches that involve segmenting a dense field motion pre-estimated [25, 26], the methods working on the estimation and simultaneous segmentation of movement to get more accurate estimate of the optical flow [27, 28], as well as approaches offering the direct parametric segmentation without estimation of motion field [29, 30] and the segmentation approaches by merging the elementary regions supplied by an initial spatial segmentation [31].

### B. Sequential Methods Working on a Pair of Images

This set includes methods which brought into play the movements can be estimated both between consecutive images [32], than between a fixed reference image and images that can be further apart [33].

### C. Method Working on a Group of Remote Images

Unlike the mentioned methods mentioned above which are interested in the estimated motion between two consecutive frames. Wills *et al.* [34] have developed a method of segmentation of two remote images for which the optical flow estimation is not efficient because these movements are relatively large. Basing on the correspondence of the interest points between two remote images.

## V. CONCLUSION

We are concentrated on three axes in this article:

Motion Detection, Motion Estimation and Segmentation based on movement.

For each of these axes we have presented the various developed methods. The classification of the motion detection methods is based on the modeling of the background. We also discussed the different techniques of motion estimation and finally the classification of segmentation approaches based on the movement is based on the number of images of the sequence to compare and the

temporal distance between them.

From the various methods mentioned above, we found that some of the solutions proposed improvements accomplished by other approaches and each as limitations. The methods developed in this article are partial; they provide only part of the solution of problems in the field of detection, motion estimation and segmentation based on movement such as occlusions, illumination changes, environment, etc. There is no general method that processes and gives a complete solution.

#### REFERENCES

- [1] S. Harass, "Image analysis video surveillance embedded in transit vehicles," Ph.D. thesis, *Polytechnic Institute of Grenoble*, 2006.
- [2] M. Brulin, H. Nicolas, and C. Maillot, "Analysis of road traffic in the context of video surveillance," *IEEE Sciences of Electronics, Technologies of Information and Telecommunication (SETIT)*, Tunisia, 2012.
- [3] D.-N. Truong C. Achard, and L. Khoudour, "Detection and re-identification of people in a network of cameras," presented at 17th Congress AFRIF French-AFIA, Pattern Recognition and Artificial Intelligence, France, 2010.
- [4] F. Zaklouta, "Object recognition for multiclass applications driving assistance and video surveillance," PhD thesis, 2011.
- [5] D. Kachi, "Scene analysis," University of Picardie Jules Verne, 2010.
- [6] N. Verbeke, "Tracking of moving objects in video sequences," Ph.D. thesis, University of Paris Descartes, 2007.
- [7] R. Jain and H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 206–214, 1979.
- [8] Y. F. Ma and H. J. Zhang, "Detecting motion object by spatiotemporal entropy," *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 265–268, Tokyo, Japan, 2001.
- [9] A. Bugeau, "Detection and tracking of moving objects in complex scenes: Application to the monitoring of drivers," PhD thesis, 2007.
- [10] T. Yang, S. Z. Li, Q. Pan, and J. Li, "Real-time and accurate segmentation of moving objects in dynamic scene," *ACM 2nd Int. Workshop on Video Surveillance & Sensor Networks (VSSN)*, pp. 136-143, New York (NY), USA, 2004.
- [11] V. Veilleux-Gaboury, "Motion Detection," University of Sherbrook, 2007.
- [12] J. Rittscher, J. Kato, S. Joga, and A. Blake, "A probabilistic background model for tracking," in *Proc. 6th European Conf. on Computer Vision (ECCV)*, vol. 2, pp. 336–350, Dublin, Ireland, 2000.
- [13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2246–2252, Fort Collins (CO), USA, 1999.
- [14] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. M. Buhmann, "Topology free hidden Markov models: Application to background modeling," in *Proc. 8th IEEE Int. Conf. on Computer Vision (ICCV)*, vol. 1, pp. 294–301, Vancouver, Canada, 2001.
- [15] N. Oliver, B. Rosario, and A. Pentland, "Bayesian computer vision system for modeling human interactions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000.
- [16] R. Sébastien, "Vision 3D," Université de Montréal (Canada).
- [17] N. Ancona and T. Poggio, "Optical flow 1d correlation Application to a simple time-to-crash detector," *Int. Journal on Computer Vision*, vol. 14, no. 2, pp. 131-146, Mars 1995.
- [18] S. M Haynes and R. Jain., "Detection of moving edges," *Computer Vision Graphics and Image Processing*, vol. 21, pp. 345-367, 1983.
- [19] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of the motion," *Journal of the optical society of America*, vol. 2, no. 2, pp. 284-299, 1985.
- [20] D. Berziat, "Medical image processing detection estimation and motion analysis," 2007.
- [21] B. K. P. Horn and B. G. Schunk, "Determining optical flow," *Artificial Intelligence*, 1981.
- [22] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *IJCAI*, pp. 674-679, Vancouver, Canada, 1981.
- [23] P. Brault, "Mouvement estimation and image segmentation," University of Paris Sud, 2005.
- [24] M. Fradet, "Contribution to the segmentation of image sequences in the sense of movement in a semi-automatic," University of Rennes, 2010
- [25] Y. A. Wang and E. H. Adelson, "Representing Moving Images with Layers," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 625-638, September 1994.
- [26] S. Ayer and S. Harpreet, "Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding," in *Proc. Int. Conf. Computer Vision (ICCV)*, pp. 777-784, 1995.
- [27] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise- Smooth Flow Fields," *Computer Vision and Image Smooth Flow Fields. Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75-104, 1996.
- [28] A. Jepson and M. Black, "Mixture models for optical flow computation" Technical Report RBCV-TR-93-44, Department of Computer Science, university of Toronto, April 1993.
- [29] J. M. Odobez and P. Boutheymy, "Direct incremental model-based image motion segmentation for video analysis," *Signal Processing*, vol. 66, no. 2, pp. 143-155, 1998.
- [30] D. Cremers and S. Soatto, "Motion Competition A Variational Approach to Piecewise Parametric Motion Segmentation," *Int. J. Computer Vision (IJCV)*, vol. 62, no. 3, pp. 249-265, Mai 2005.
- [31] F. Moscheni, S. Bhattacharjee and M. Kunt, "Spatiotemporal Segmentation Based on Region Merging," *IEEE Trans. Pattern Anal. Machine Intell. (PAMI)*, vol. 20, no. 9, pp. 897-915, 1998.
- [32] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Proc. Int. Conf. Computer Vision (ICCV)*, pp. 1154-1160, 1998.
- [33] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion Magnification," *ACM Trans. on Graphics (SIGGRAPH)*, 2005.
- [34] J. Wills, S. Agarwal, and S. Belongie, "What Went Where," *Conf. Computer Vision Pattern Rec. (CVPR)*, vol. 1, pp. 37-454, Madison, WI, Juin 2003.

**Amara Kahina** was born in Algiers in October 1988. she obtained her master's degree in control and robotics at the University of Science and Technology USTHB (Algiers) in 2011, and she is working on her PhD at USTHB in computer vision.

**Djekoune A. Oualid** was born in Algiers, Algeria, in October 1969. He received his Ph.D. degree in process control and robotics from the electronic institute of USTHB, Algiers, in 2010. He is the author of numerous publications for conferences proceedings and journals. He has been active in research on perception, computer vision, pattern recognition and robotics. Now, he works as researcher in the vision team of the Robotics and Industrial Automation Division in high research center in Algiers.

**Belhocine Mahmoud** gained his doctor at in robotics in 2006 from USTHB, Algiers (Algeria). He is now head of the Vision Team of the robotics division of the CDTA (Centre de Développement des Technologies Avancées) in Algiers. His fields of interest are computer vision, pattern recognition, augmented reality and visual control.

**Achour Nouara** is a professor at University of USTHB at Algiers. She received a degree in electronics engineering from the ENP (ecole nationale polytechnique) at Algiers in 1983, followed by the magister degree in electronics applied to control processing in 1992 and a Ph.D degree in motion planning in 2004 from the USTHB University. She is the director of research at LRPE (Laboratoire de Robotique, Parallélisme et Electro-énergétique) at USTHB. Her main research activities are in the areas of robotics, autonomous robots, artificial intelligence, motion planning and haptics.