Classifications of images for the recognition of people's behaviors by SIFT and SVM

Henni Sid Ahmed *
Department of Electronic Engineering
University of Oran Algeria laboratory LSSD USTO

Belbachir Mohamed Faouzi
Department of Electronic Engineering
University of Oran Algeria laboratory LSSD USTO

Abstract— Bbehavior recognition has been studied for realizing drivers assisting system and automated navigation and is an important studied field in the intelligent Building. In this paper, a recognition method of behavior recognition separated from real image was studied. Images were divided into several categories according to the actual weather, distance and angle of view etc. SIFT was firstly used to detect keypoints and describe them because the SIFT (Scale Invariant Feature Transform) features were invariant to image scale and rotation and were robust to changes in the viewpoint and illumination. My goal is to develop a robust and reliable system which is composed of two fixed cameras in every room of intelligent building which are connected to a computer for acquisition of video sequences, with a program using these video sequences as inputs, we use SIFT represented different images of video sequences, and SVM (support vector machine) Lights as a programming tool for classification of images in order to classify people's behaviors in the intelligent building in order to give maximum comfort with optimized energy consumption.

Keywords— SIFT; image analysis; people behavior; recognition

I. INTRODUCTION

Intelligent building sector is now one of the sectors that consume more energy than the transport and industry. In France for example, the building sector is responsible for 21% of CO_2 emissions and 43% of the total energy consumption [1]. Industrialized countries committed to reduce their serre effect gas emissions. The Building sector offers great opportunities for saving energy; Several solutions are currently proposed to achieve these savings. There is first of all the decentralized production of energy from renewable energy, passive solutions for saving energy (insulation etc),Then the Solutions of the active management of energy consumption. To enable the development of these solutions, it is essential to have reliable information on the occupation of the buildings, that is to say, we need to know the behavior of people in intelligent building for better management and optimization of energy consumption. The recognition of people's behavior provide information that will enable the system to better manage and optimize energy consumption in the intelligent building. The recognition of human activities is widely studied in recent years by the scientific community in computer vision. Reference articles [2, 3, 4] describe carefully the methods used in the state of the art. Activity Recognition is, from low-level information such as the numerical value of a set of pixels, to obtain a semantic representation is a natural language scene. The process of recognition of activity can be considered as a classification problem, where the various Representations of activities and recognition techniques, are involved, it is a very complex problem. That can be detected independently of each other; we must then analyze their temporal distribution to get to the scenario or activity. Several simplifying assumptions and methods have been proposed in the literature to solve the problem of the change of point of view. Firstly, some methods use multiple cameras [5, 6]. This assumption greatly simplifies the problem but can not be used in many situations where only a monocular system is used [7]. With a single camera, some authors have proposed the use of a camera model (model ane [8] or projective [9] for example). Other authors are based on the équipolaire geometry between the same pose under two different angles [10, 11]. For example, Rao et al. [12] Syeda-Mahmood et al. [13] compute the correspondence between two actions in assessing the fundamental matrix with a few selected points. Representations can be used explicitly formulated or determined by learning. Some methods use physical action such as a recurring character. [14] Many methods use directly the movement [15, 16] where are based on images from the pre-planning to build eg Motion Energy Images (MEI), the Motion History Images (MHI) [17] or more recently Motion History Volumes (MHV) [5]. The behavior analysis is usually done in two steps: description and recognition of actions. The first step is to define a model that describes each relevant action in our application context. Then there are two possibilities. First, there is a training phase using labeled data and then recognize the new data, based on this training. The methods used are hidden Markov models, neural networks, the support vector machines (SVM) ,etc [18]. In this paper, we present a system for classifications of images For The recognition of people's behaviors by SIFT and SVM so that, this system communicate the behavior of individuals to another management system, as it will give maximum comfort to people with the optimization of energy consumption in real time. Our system is composed of several fixed camera installed in the environment where we would like to know the individuals activities, exactly two fixed camera in every room of intelligent building, which are connected to a computer for the acquisition of video sequences in real time, and a program for the classification of people's behavior in real time to optimize power consumption in the intelligent building (lighting, heating, etc..) with maximum comfort. We selected 11 behaviors to know people's behaviors and each behavior is represented by a set of images, the fixed cameras are with great qualities which gave us images with high precision, we used the SIFT (Scale Invariant Feature Transform) represented for the different images obtained from video sequences, and SVM Light (Support Vector Machine) for implementation because it is a very powerful programming tool which is with a very low calculation and learning times with a very such as learning times our obtained results show the advantages of the use of desriptor SIFT in the field of The recognition of people's behaviors by SIFT and SVM

II. EXPERIENCES

We have used the same data base that we have realised in the lab team Multicom which is the building CTL (center technology and software) at the University Joseph Fourier of Grenoble France in 2010 for the realisation a international publication, that have published in 2013[18]. Each piece of intelligent building is equipped with two cameras to record the behaviors of people; we choose 11 behaviors that cover most of the behaviors of individuals in a smart building at maximum energy consumption. We extracts e the various images from video sequences that define the11 behaviors [18], and we built our database which is the set of still images, then we use SIFT for parameterization images, such that each image is represented by a vector of the same size elements 128.For implementation, we apply using the SVM (support vector machine) learning method on these data in order to reduce the error recognition system and a method of classification to determine the behavior of the people in this intelligent building. The choice of scenarios is a very important step as it is necessary to choose scenarios that cover the behavior of people in an intelligent building. After a research on the different work already done in this area we have chosen the relevant scenarios that allow us to determine the behavior of 11 people in an intelligent building [18], which are: 1) I am sitting on the bed and I am reading a document, 2) I do the gym in the room, 3) I do the washing up, 4) I walk from one room to another in DOMUS (Intelligent Building) , 5) I eat on the kitchen table, 6) I make coffee and I drink it standing 7) I'm sitting in the office and I work at the computer, 8) I'm lying in bed, 9) I'm sitting on the bed and I watch TV, 10) I'm sitting in the office listening to music 11) I entere DOMUS (Intelligent Building) and I settle myself. we took into consideration the lighting and its variation is a very important factor to be taken into account and the change of person who play scenarios, as we changed the degree of opening and closing windows and flaps and the degree of lighting of the lamps and to repeat the same scenario by the same person, then by different persons in order to obtain a robust recognition system.

SVM will generate 11 models in the learning phase or training phase, so that each model corresponds to a behavior Figure 1, for the classification of behavior, the SVM use the 11modeles obtained in the learning phase, we use the models and the images that correspond to each behavior. The classification of the behaviors is according to the classification rate it means the behavior that we want is the one that have the highest rate of classification Figure 2

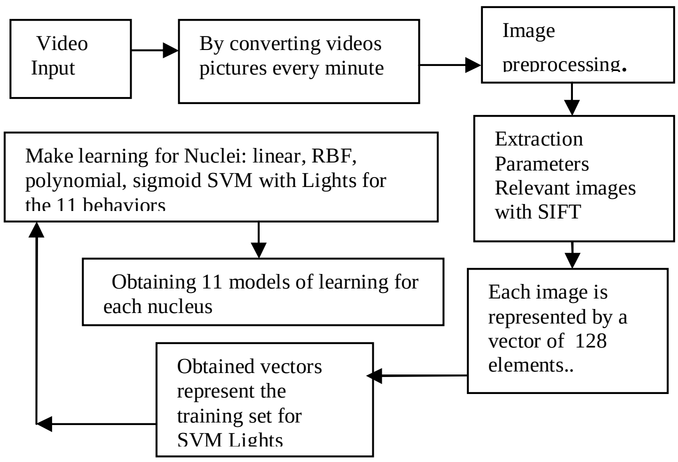

Fig. 1 Learning behaviors by SVM Lights

We run tests on the data to understand system behavior in various situations. As the system generates a report for each person and for each frame of the video, we compare these results with the ground truth, we then calculate the percentage of correct states (see tables below).

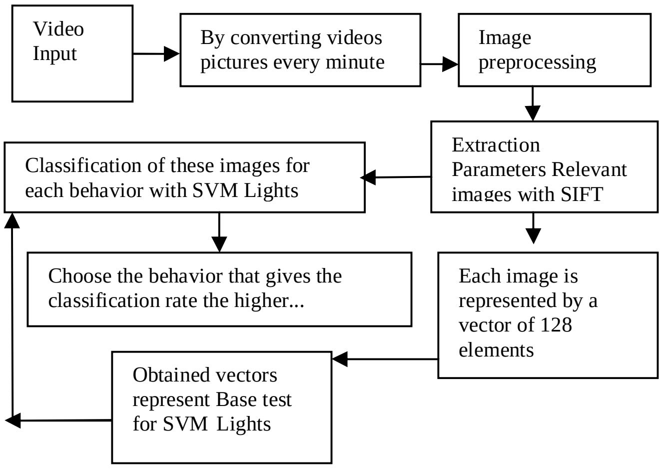

Fig. 2 Behavior classification by SVM Lights

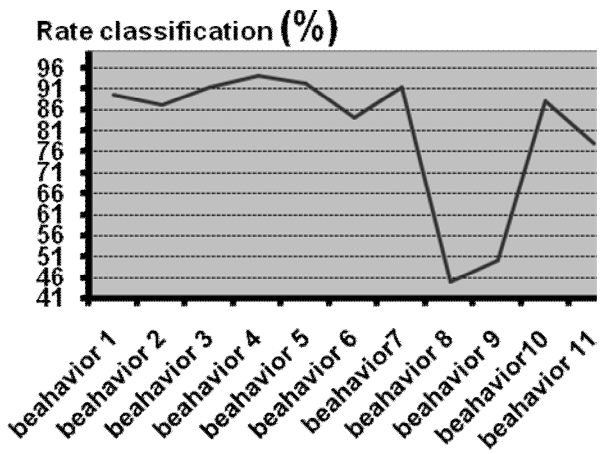

Fig. 3 Showing the percentage of correctly Detected behavior for the linear kernel

TABLE. I. Painting of the recall and precision for the Detection of behavior for the linear kernel

behavior classified	**precision linear kernel**	**reminder for linear kernel**
behavior 1	44.45%	39.15%
behavior 2	30%	44.21%
behavior 3	98%	25.11%
behavior 4	87.27%	63.52%
behavior 5	53.48%	64.92%
behavior 6	31.12%	42%
behavior 7	57.54%	84.90%
behavior 8	67.90%	34.95%
behavior 9	56.91%	87%
behavior 10	62.52%	21.36%
behavior 11	26%	74.68%

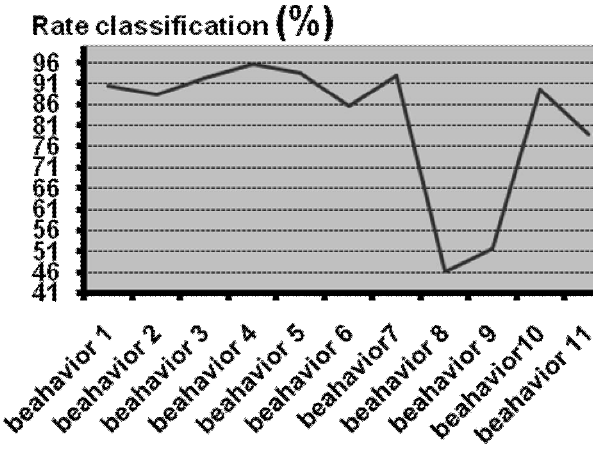

Fig. 4 Showing the percentage of correctly Detected behavior for the RBF kernel

TABLE. II-- .Painting of the recall and precision for the Detection of behavior for the RBF kernel

behavior classified	**precision RBF kernel**	**reminder for RBF kernel**
behavior 1	45.68%	40.95%
behavior 2	31.25%	45.12%
behavior 3	100.00%	26.52%
behavior 4	88.27%	67.82%
behavior 5	55.88%	66.42%
behavior 6	32.42%	43.74%
behavior 7	58.52%	86.60%
behavior 8	69.29%	36.79%
behavior 9	58.91%	88.45%
behavior 10	63.54%	22.56%
behavior 11	27.48%	76.85%

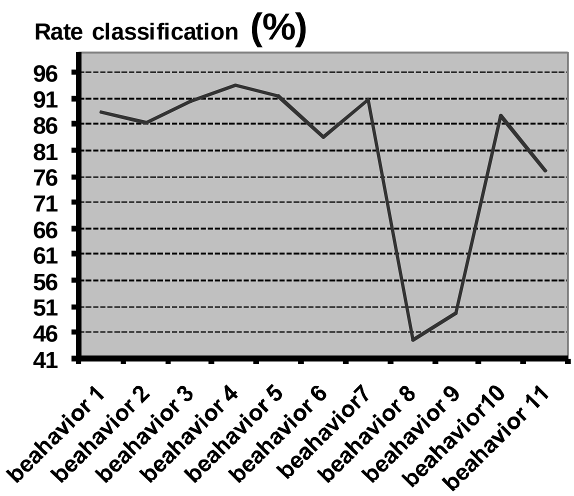

Fig. 5 Showing the percentage of correctly Detected behavior Polynomial kernel

Table. III

. Painting of the recall and precision for the Detection of behaviors for Polynomial kernel

behavior classified	**precision Polynomial kernel**	**reminder for polynomial kernel**
behavior 1	63.12%	32.88%
behavior 2	30.95%	42.26%
behavior 3	97.26%	24.98%
behavior 4	97.56%	27.91%
behavior 5	62.95%	63.12%
behavior 6	30.98%	42.11%
behavior 7	40.14%	85.20%
behavior 8	32.11%	42.45%
behavior 9	71.57%	19.51%
behavior 10	69.25%	36.15%
behavior 11	29.34%	42.15%

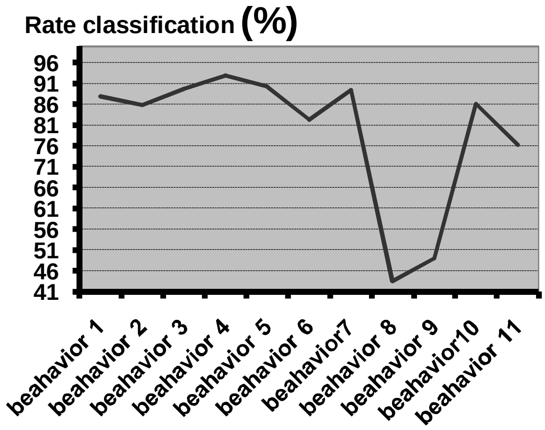

Fig 6 Painting of the percentage of correctly Detected behavior sigmoid kernel

TABLE VI -- Painting of the recall and precision Detection of behaviors for Sigmoid kernel

behavior classified	**precision Sigmoid kernel**	**reminder for Sigmoid kernel**
behavior 1	46.15%	41.10%
behavior 2	32%	44.22%
behavior 3	100.00%	27.91%
behavior 4	100.00%	28.22%
behavior 5	65.75%	65.75%
behavior 6	31.97%	44.83%
behavior 7	59.83%	87.50%
behavior 8	32.41%	45.05%
behavior 9	100.00%	45.60%
behavior 10	66.67%	21.05%
behavior 11	36.69%	51.98%

According to the results in the tables (1,2,3,4) and in the figures (3,4,5,6) above, we can make the following deductions: 1) we notice that the behaviors 8 and 9 have the less classification rate, because in these two behaviors there are a lot of a dark images which are not well presented by SIFT, 2) the beahaviors that have rapid movements of the person, have the less classification rate he 3) the results presented above show a very satisfactory performance especially for the RBF kernel because in addition to the highest recognition rate the response time is the rapid compared to other nuclei especially the time of learning, 4) it is more difficult to detect certain behaviors than others, and this is due to the nature of the behavior detected, the behavior of going into the domus and settle in, is more difficult to be detected than the behavior of doing the washing up in the kitchen.

III. CONCLUSIONS

We present in this paper a system for classifying images in order to classify people's behaviors in the intelligent building by SIFT and SVM Lights. This system allows us to characterize the activity of people in a room. This information will be useful to the management system of the building regulate the consumption of electrical energy in order to optimize (lighting, heating, etc...). SIFT is good method for the parameterization of images, robust to image rotation, scale, intensity change, and to moderate affine transformation, dark images which are not well presented by SIFT. Our classification system was evaluated on a broad base of videos to get a good learning. We used several nuclei in the learning phase to select the best model, the experimental results are satisfactory one the used method which allows us to say that the classifications of 11 behaviors in real time gave good results using the RBF nucleu SVM Lights and for learning and SIFT for the parameterization of images that is to say the images conversion into vectors of the same size.. Future work will improve these results with the acquisition of new data such as speech for the recognition the behaviors .

REFERENCES

[1] Efficacité énergétique des bâtiments 2007-2008, ‘‘rapport du ministère de l'écologie du développement et de l'aménagement durable’’, 2007.
[2] J. K. Aggarwal and Q. Cai, ‘‘Human motion analysis : a review’’,Computer Vision and Image Understanding, 73:90_102, 1999
[3] D. M. Gavrila, ‘‘The visual analysis of human movement : a survey’’, Computer Vision and Image Understanding, 73 :82_98, 1999
[4] W. Hu, T. Tan, L. Wang, and S. Maybank, ‘‘A survey on visual surveillance of object motion and behaviors’’, Systems, Man, and Cybernetics, Part C :Applications and Reviews,,34(3) :334_352, 2004
[5] R. Kehl, M. Bray, and L.Van Gool, ‘‘Full body tracking from multiple views using stochastic sampling’’, interantional conference on Computer Vision and Pattern Recognition, 2 :129_136, 2005.
[6] D. Weinland, R. Ronfard, and E. Boyer, ‘‘Free viewpoint action recognition using motion history volumes’’, Computer Vision and Image Understanding , 104(2) :249_257, 2006.
[7] F. Lv and R. Nevatia, ‘‘Single view human action recognition using key pose matching and viterbi path searching’’, international conference on Computer Vision and Pattern Recognitio, pages 1_8, 2007
[8] C. Rao, A. Yilmaz, and M. Shah, ‘‘View-invariant representation and recognition of actions’’, International Journal of Computer Vision, 50(2) :203_226, 2002.
[9] V.Parameswaran and R. Chellappa, ‘‘View invariance for human action recognition’’, International Journal of Computer Vision, 66(1) :83_101, 2006.
[10] A. Gritai, Y. Sheikh, and M. Shah, ‘‘On the use of anthropometry in the invariant analysis of human actions’’, International Conference on Pattern Recognition, 2 :923_926, 2004.
[11] A. Yilmaza and M. Shah, ‘‘Matching actions in presence of camera motion’’, Computer Vision and Image Understanding, 104(2) :221_231, 2006.
[12] C. Rao, A. Gritai, M.Shah, and T. Syeda-Mahmood, ‘‘View-invariant alignment and matching of video sequences’’, International Conference on Computer Vision, 2 :939_945, 2003.
[13] T. Syeda-Mahmood, A. Vasilescu, and S. Sethi, ‘‘Recognizing action events from multiple viewpoints’’, Detection and Recognition of Events in Video Workshop, pages 64_72, 2001.
[14] Qiang He and C. Debrunner, ‘‘Individual recognition from periodic activity using hidden markov models’’, Human Motion Workshop, pages 47_52, 2000.
[15] A.A. Efros, A.C. Berg, G. Mori, and J. Malik, ‘‘Recognizing action at a distance’’, International Conference on Computer Vision, 2 :726_733, 2003.
[16] R. Cutler and M. Turk, ‘‘View-based interpretation of real-time optical ow for gesture recognition’’, International Conference on Automatic Face and Gesture Recognition, pages 416_421, 1998.
[17] J.W. Davis and A.F. Bobick, ‘‘The representation and recognition of action using temporal templates’’, International conference on Computer Vision and Pattern Recognition, pages 928_934, 1997.
[18] Henni Sid Ahmed, Belbachir Mohamed Faouzi, Jean Caelen , ‘‘ Detection and classification of the behavior of people in an intelligent building by camera’’, International Journal On Smart Sensing And Intelligent Systems vol 6, no 4, September 2013

