exeley INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Article | DOI: 10.21307/ijssis-2020-034 Issue 1 | Vol. 13 (2020)

The command of comfort in an intelligent building by speech classification and image classification for energy optimization

Henni Sid Ahmed[1,*] and Jean Caelen[2]

[1]Université Abdelhamid Ibn Badis of Mostaganem Algeia, Faculty of Sciences and Technology Departement Genie Electrical, Laboratory LSS, Route Nationale N°11, Kharrouba, Mostaganemm, 27000, Algeria.

[2]Université Joseph Fourier, Grenoble, F, Institut Carnot LSI, Allée de Palestine, 38610 Gières, France.

*E-mail: sidhenni@yahoo.fr

This paper was edited by Subhas Chandra Mukhopadhyay.

Received for publication September 29, 2020.

Abstract

Comfort command is a solution to optimize energy in an intelligent building. Our goal is to achieve an optimum and robust system, for the command of comfort. Speech and image classification are considered to be two systems that have worked well for comfort command, but both systems have drawbacks. After several studies, we noted that the inconveniences of the speech classification are solved by adding in parallel a classification of the image and the same thing for the disadvantages of the classification of images that are resolved by adding in parallel a system of speech classification. It means that these two systems work at the same time to achieve a robust and optimum system for comfort command.

Keywords

Comfort, Intelligent building, Speech processing, Image processing, SVM, Classification.

It is very difficult to define comfort in the context of building standards of intelligent buildings. Approaches that define thermal comfort give two definitions of comfort. The first defined by Fanger (2009), who defines comfort as a state of neutrality, means a state of mind expressing the satisfaction of its environment, in this case the person cannot say if he wants to increase heating (having warmer) or air conditioning (getting colder). The second, defined by Givoni and Izard (1978) defines comfort as the conditions for which the self-regulating mechanisms of the body are at a minimum level of activity. The intelligent building has a very important role, including measuring comfort. Saizmaa and Kim (2008) analyze the issue of smart building design. They developed the notion that associates the intelligent building with psychological, emotional, physical, social, and spiritual appropriations. Ekambi Schmidt (1972) was interested in the relationship between smart buildings and the people who occupy them, and the representation that these people construct of their place of life. Moser (2009) describes this relationship as the basis in home, a physical and abstract space in which the intelligent building must be able to express and identify itself. There are a lot of projects devoted to the use of speech-recognition technologies, such as Companions (Cavazza et al., 2010), Companion Able (Rougui et al., 2009), Aladin (Gemmeke et al., 2013), or Pipin (Casanueva et al., 2014). However, there are still many questions about the real impact both in terms of assistance and installation cost, despite real evaluations in test apartments with Pers projects (Hamill et al., 2009), Dirha (Ravanelli and Omologo, 2014). Our contribution in this field of research is the realization of a robust and optimum system for the control of comfort in an intelligent building. The speech classification system is a very robust system that gives very good results for the control of comfort in an intelligent building. Its advantage over other existing solutions is that it gives the comfort desired by the person at all times and with great precision, for example, if we compare it with another system that uses the behavior of people by camera (Ahmed et al., 2013), to give them the desired comfort, the latter is not as precise as the speech classification system, because a behavior does not reflect 100%

the desired comfort, but the person's speech that reflects 100% the desired comfort is the reason that prompted us to use the classification of speech to command comfort. But after using this system, we noticed that it has some faults, such as in some situations, it will not be able to command the comfort, and these situations are:

1. If the person is sick to a point where he will not be able to speak, or if his voice has changed a lot, then in this case, the system will not be able to command the comfort and the energy will not be optimized.
2. If the person is asleep leaving the lights on, the television, or music on, then in this case, the energy will not be optimized, because these equipments will not be turned off while that person is asleep.
3. If the person leaves the room, then the comfort management system always continues to give comfort because it is not informing of the absence of the person in the room, and therefore the energy is not optimized.

The solution to these defects is to use the speech classification system with an image classification system, to add another image classification system in parallel with the speech classification system, and use these two systems at the same time for comfort command.

This system of classification of images will allow to solve the defects of the system of classification of speech. We have noticed that to solve these defects, the image classification system must classify three images:

1. The first image expresses the presence of the person.
2. The second image expresses a sleeping person.
3. The third image expresses the absence of the person.

When this system detects the presence of a person, then the intelligent building comfort management system gives optimum comfort to the person. We need this system in the case of a sick person who cannot speak so that he cannot command the comfort. If the system does not detect the presence, the intelligent building comfort management system turns off all lights and other equipment in that room, and lowers the heating or the air conditioner.

In the event that the system detects that the person is asleep, then the intelligent building comfort management system turns off all lights and other equipment in that room. The disadvantage of the image classification system is that it requires lighting, and therefore in the dark, the person will not be able to have his comfort; in addition, there are people who prefer to stay in the dark and to have their comfort. The solution to this inconvenience is the speech classification system, and therefore we notice that these two systems complement each other.

In the first system of speech classification, an experiment will be implemented using an input database, the vocabulary used in this database is made up of 20 words in the English language. The classification will be of global type, the voice signal will be represented using the MFCC (Mel Frequency Cepstral Coefficients) coefficients to form the input vector of the classification system, and the SVMs will be used for the learning and the voice recognition phase. In the second image classification system, the SIFT (Scale Invariant Feature Transform) method is used for the preparameterization of the images, and SVM will be used for the learning and the image recognition phase. We opted for SVM (Support Vector Machines) because of their good generalization capacities in many problems, but also for their good results on sets of small training data (unlike neural networks that require a larger database). They have shown in recent years their power, especially their computing time, which is very satisfactory compared to other techniques (Joachims, 1999).

In order to improve performance, a study on several nuclei (Linear, RBF, Polynomial, and Sigmoid) was carried out in the learning phase, taking the nucleus, which gives the best results.

Support vector machine (SVM)

SVM is a classification method which was introduced by V. Vapnik (1995). Support vector machines are known as support vector networks, which is a supervised learning used in machine learning (Srinivasan and Rajakumar, 2017). SVM is a classification technique based on kernel methods that has been proved very effective in solving complex classification problems in many different application domains (Dixon and Candade, 2008). This method is characterized by solid theoretical foundations which make its success. There is indeed a direct link between the theory of statistical learning and the SVM learning algorithm. The drawback of most ML (Machine Learning) techniques is that they have a large number of learning parameters to be set by the user (structure of a neural network, gradient update coefficient, etc.). In addition, with these methods, the number of

parameters to be calculated by the learning algorithm is linear, even exponential, with the dimension of the input space.

An SVM teaches a separator. This brings the problem back to the definition of a separator. Let us give a finite set of vectors of Rn, separated into two groups, that is to say into two classes. Membership in one class or another is defined by a label, associated with each of the vectors, on which is written 'class 1' or 'class 2'. Finding a separator is like building a function, which takes a vector from our set, and can tell which class it is. SVMs are a solution to this problem, as would be a simple rote learning of the classes associated with the vectors of our set.

For the learning, the principle of SVM, based on the theorem of statistical learning, consists first of all in transforming the inputs into a space, thanks to the kernel, we take a linear decision border in this space (Karush, 1939). Support vector machines are known as Support vector networks, which is a supervised learning used in the learning (Srinivasan and Rajakumar, 2017). Among the SVM models, there are linearly separable cases and non-linearly separable cases. The implementation of SVM classifiers in the context of the linearly separable case is simple. In most real problems, there is no possible linear separation between data, the classifier (hyperplane separator) cannot be used because it only works if the classes of training data are linearly separable. To overcome the drawbacks of non-linearly separable cases, the idea is to use the kernel function which consists in changing the data space, by performing a transformation of the representation space of the input data into one more space. Large dimension (possibly of infinite dimension), in which these input data are linearly separable. In practice, a few families of configurable kernel functions are known, and it is up to the SVM user to carry out tests to determine which one is best suited for his application. We can cite the following examples of nuclei: polynomial, Gaussian, sigmoid, and RBF (radial basic function).

In most real problems, there is no possible linear separation between the data of the input data base, which is the case for our data base. It is for this reason that we use the kernel functions so that our data is separable. For the implementation, we use several kernels in order to take the best models which give us the highest classification rate.

For the recognition SVM (Support Vector Machine) aims to find the two best hyperlinks for separating word or image into two classes. To make a classification you must: determine an appropriate classification system, select training samples, image pre-processing, extract features, select fitting classification approaches, post-classification processing and accuracy assessment (Prasad et al., 2017). The classification of a word or image is given by its position in relation to the best hyperplane (model or classifier) found in the learning phase. We used the so-called 'one against all' approach for classification, which successively compares one class to all the others using a binary classifier.

The automatic speech recognition system

Experience

We conducted experiments in the intelligent building at the LIG laboratory, located at the University of Grenoble, France, specifically at the CTL (Software Development Center). It is a fully functional 35-m^2 apartment composed of a kitchen, a bedroom, a bathroom, and an office.

Principle

The Word recognition by SVM in a natural environment is a global classification. It has four steps:

1. Acquisition of the speech signal and formation of the input corpus.
2. Pretreatment and parameterization.
3. SVM for the learning design.
4. SVM for recognition.

Acquisition

The acquisition of the speech signal was carried out in the intelligent building, such as the microphones, which are fixed in each room, as well as the kitchen and the hallway. The speech signal will be picked up by these different microphones, then it will be recorded and stored in a database. The apartment is also equipped with more than 150 home automation sensors connected by a home automation network. These sensors used are of the KNX type, these sensors allow the control of lighting, temperature, shutters, security, heating, ventilation, air conditioning, water, and energy, but in our case, we do not need these sensors, because we would like to control the comfort by the classification of the speech and the image.

Database input

As we have already mentioned, comfort is divided into three types: thermal comfort, acoustic comfort

tick, and the visual comfort. So, to command comfort in the smart building, just order these three types of comfort.

So the vocabulary we used consisted of 20 words in English that control these three types of comfort. These words are W1: light on, W2: light off, W3: Light up, W4: Light down, W5: Window on, W6: Window off, W7: Window up, W8: Window down, W9: music on, W10: music off, W11: Music up, W12: Music down, W13: Heater on, W14: Heater off, W15: Heater up, W16: Heater down, W17: Could on, W18: Could off, W19: Could up, W20: Could down.

Each word is pronounced three times by each speaker, so that the person pronounces the same word differently, for example when he is happy, sad, and tired. We used 24 speakers (8 children, 8 men, and 8 women, whose age varies between 17 and 65, four children between 5 years and 16 years). So our input corpus will contain 4,320 ($24 \times 3 \times 3 \times 20 = 4{,}320$) words uttered by these 24 speakers, who will be recorded and stored in a file called database. In total, 70% of the input database will be used for the learning phase, and the remaining 30% will be used for the recognition phase. So, 3,024 pronounced words will be used for learning, and 1,296 pronounced words will be used for recognition.

Pretreatment

The recorded words are tainted, at their beginning and end, by silence. The latter is to be deleted in order to keep only the significant information that represents the word itself and this, thanks to a delimitation procedure that determines the beginning and the end of the words. Since the signal of the speech is not stationary, then we have to break the signal into a set of windows, so that it is stationary in each window. We opted for the Hamming window, because it gives good results. The speech signal is decomposed into a hamming window of 23 ms (256 speech samples), by performing a recovery of 6 ms not to lose the information between the two windows.

Parameterization

MFCCs are the cerebral coefficients of frequency Mel, they allow the shape of the spectrum of a signal to be described using a reduced number of coefficients.

We have used the Mel FCC (Mel Frequency Cepstral Coefficients) method for the parameterization of the speech signal because it has already been proved that Mel FCCs can completely and efficiently represent any signal in 13 coefficients, and the results obtained by the Mel FCC are very satisfactory. The MF Frequency Coefficient Cepsctrel method is based on a perception scale called Mel, which is nonlinear, and its frequency scale based on human perception. The spectrum obtained, before extracting the cepstrum, is more representative of how the brain works. This can be defined by the relation between the frequency in hertz and its correspondence in mels. Since the signal of the speech is stationary, then we have to break the signal into a set of windows, so that it is stationary in each window. We opted for the Hamming window, because it gives good results.

After having decomposed the signal into a 23-ms window, the Mel FCC method is applied, in order to extract the significant parameters of the speech signal. A vector of 13 Mel FCC coefficients would be extracted from each window; we will calculate each of its coefficients and the signal will be expressed as a linear combination of its past.

Each word will be divided into 21 segments. Each segment will undergo a 23-ms HAMMING windowing (256 samples) with a 40% overlap. From each part, 13 Mel FCC coefficients are extracted. Order 13 is sufficient to characterize a part. After concatenation of the different blocks of results, each word will be represented by 273 coefficients (21 segments coded with 13 coefficients). $W_{i\,Mel\,FCC}$ represents the word w by the Mel FCC method, and $W_{i\,j}$ the coefficients.

$$W_{i\,Mel\,FCC} = \begin{matrix} W_{1\,1} & W_{1\,2} & \cdots & W_{1\,21} \\ W_{2\,1} & W_{2\,2} & \cdots & W_{2\,21} \\ W_{3\,1} & W_{3\,2} & \cdots & W_{3\,21} \\ W_{4\,1} & W_{4\,2} & \cdots & W_{4\,21} \\ W_{5\,1} & W_{5\,2} & \cdots & W_{5\,21} \\ W_{6\,1} & W_{6\,2} & \cdots & W_{6\,21} \\ W_{7\,1} & W_{7\,2} & \cdots & W_{7\,21} \\ W_{8\,1} & W_{8\,2} & \cdots & W_{8\,21} \\ W_{9\,1} & W_{9\,2} & \cdots & W_{9\,21} \\ W_{10\,1} & W_{10\,2} & \cdots & W_{10\,21} \\ W_{11\,1} & W_{11\,2} & \cdots & W_{11\,21} \\ W_{12\,1} & W_{12\,2} & \cdots & W_{12\,21} \\ W_{13\,1} & W_{13\,2} & \cdots & W_{13\,21} \end{matrix}$$

The use of cepstral coefficients has the advantage of allowing a faster comparison between two samples. Indeed, it makes it possible to calculate a distance from a number of coefficients much lower than that of a conventional frequency spectrum, the information on the spectral envelope being compacted into the first coefficient of the cepstrum.

MFCCs take into account human perception of sensitivity to appropriate frequencies by converting conventional frequency to Mel Scale, and are suitable

for understanding humans and the frequency at which humans speak.

Example of MFCC coefficients:
–6.249e + 02
–5.834e + 02
–6.999e + 02
–7.299e + 02
8.105e – 15
4.787e + 01
MFCC coefficients = –8.105e – 15
–8.105e – 15
2.066e – 14
–8.500e + 00
2.421e – 14
4.109e – 14
3.931e – 14

If a cepstral coefficient has a positive value, it represents a sound that belongs to the low frequencies, and in this case, the majority of the spectral energy of the sound is concentrated in the low frequencies.

On the other hand, if a cepstral coefficient has a negative value, it represents a sound that belongs to the high frequencies, and in this case, the majority of the spectral energy of the sound is concentrated in the high frequencies.

The lower-order coefficients contain most of the information about the overall spectral shape of the source-filter transfer function.

- The zero-order coefficient indicates the average power of the input signal.
- The first-order coefficient represents the distribution of spectral energy between low and high frequencies.

Learning the network

In total, 70% of the words in the database represented by Mel FCC method are dedicated for learning. For each word, we have taken all the samples that represent this word that is represented by 273 coefficients.

For the implementation, we use several kernel functions to take the best models that give us the highest classification rate. Since we have 20 words to recognize, the SVM will generate 20 models (classifiers or hyperplane separators) in the learning or training phase, as each model corresponds to one word. In order to improve performance, a study on several nuclei (Linear, RBF, Polynomial, and Sigmoid) was carried out in the learning phase, taking the nucleus, which gives the best results.

Recognition

In total, 30% of the words in the database represented by Mel FCC method are dedicated for recognition. For each word, we have taken all the samples that represent this word that is represented by 273 coefficients.

As we have 20 words to recognize for the automatic speech recognition system, then the SVMs will generate 20 models (classifiers or hyperplane separators) in the learning phase, for the first system so that each model corresponds to a word. Since we have 20 classes, then we will have 20 binary classifiers, which are constructed (20 hyperplanes, or 20 problems of optimization), the *n* classifier being intended to distinguish the class of index *n*. Each element to be classified is presented to the 20 classifiers and is assigned the label of the classifier that returned the highest percentage of adhesion.

Calculation of the recognition rate (TR)

After the learning phase, we test the system with the vectors who have never served. On-line tests will be done on the network before calculating the recognition rate: see Figure 1.

Recognition rate = number of times or the word is recognized/number of attempts.

Resulte and discussion

1. $$\text{The rate of recognition} / W_i = \frac{NRW_i}{N_T W_i}, \quad (1)$$

   NRW_i: represents the number of recognized word W_i.
   $N_T W_i$: represents the total number of the word W_i in the base of tets.
   RR: represents the rate of recognition.
2. The precision for a W_i word: is the number of words correctly detected in divided by the total number of words in the entire test base that are detected and assigned to the word W_i: (https://fr.wikipedia.org/wiki/Pr%C3%A9precision_et_rappel).
   $N_{cd} W_i$: the number of words correctly detected for the word W_i.
   $N_d W_i$: the total number of words detected for the word W_i in the test base
   So:

$$\text{Precision} / c_i = \frac{N_{cd} W_i}{N_d I_k W_i}. \quad (2)$$

3. The reminder for a W_i word; this is the number of correctly detected words for the W_i word divided by the total number of detected words that belong to the W_i word (https://fr.wikipedia.org/wiki/Pr%C3%A9precision_et_rappel).
N_{cd} W_i: the number of words correctly detected for the word W_i.
N_T W_i: the total number of detected words that belong to the word W_i.
So:

$$\text{Reminder} / c_i = \frac{N_{cd} W_i}{N_T W_i}. \quad (3)$$

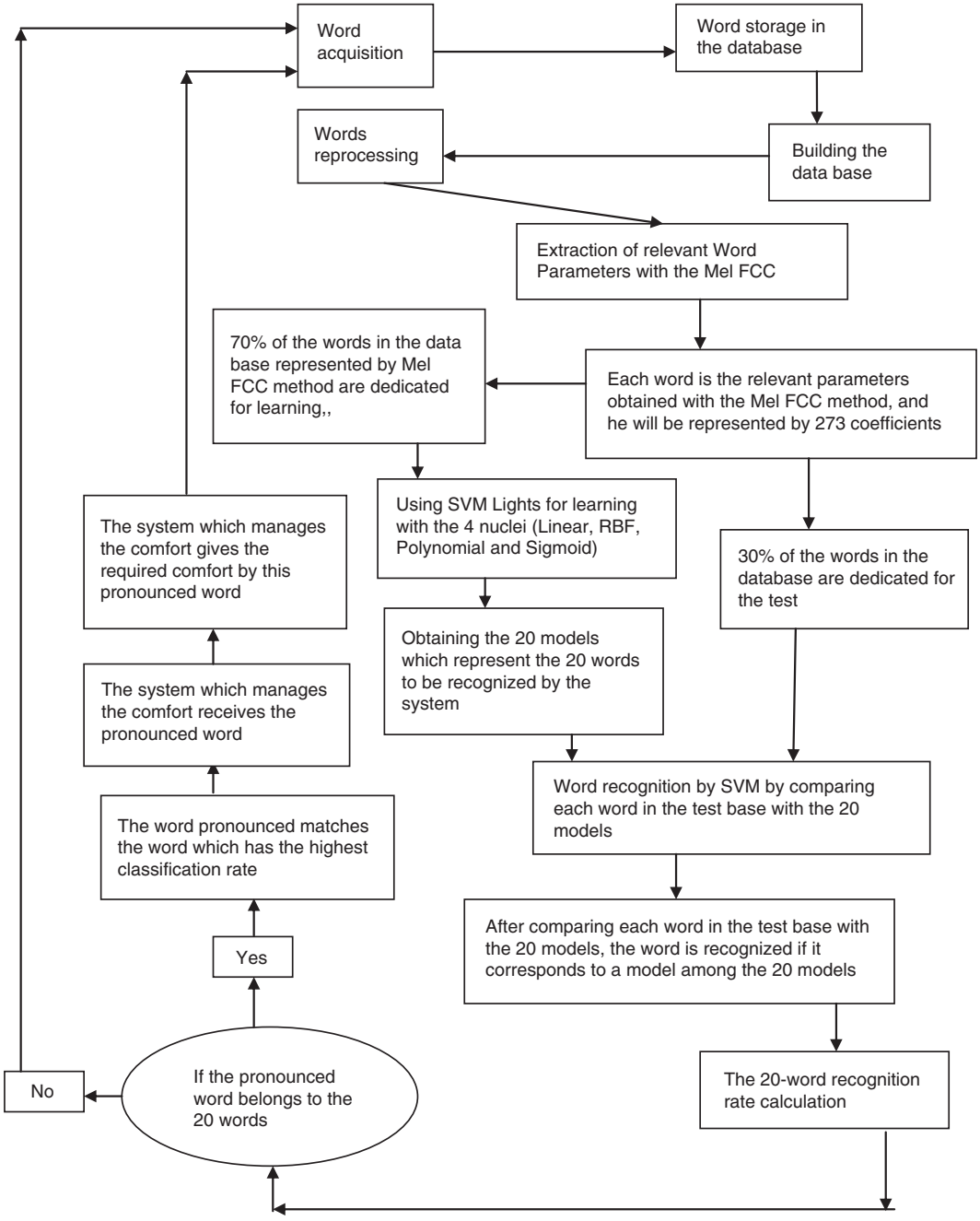


Figure 1: The diagram of the automatic speech recognition system by MFCC and SVM.

As long as the values of the recognition rate, precision, and recall are high, we can guarantee reliability and accuracy of the classification system.

The classification is done by the SVM. To improve this classification, we used four nuclei, linear, polynomial, RBF (radial basic function), and sigmoid.

The results of the experiments are given in Figures 2-9 and in Tables 1-4.

Classification with SVM using the linear kernel nucleus Figures 2 and 3 and Table 1.

Classification with SVM using RBF (radial basic function) nucleus Figures 4 and 5 and Table 2.

Classification with SVM using the polynomial kernel nucleus Figures 6 and 7 and Table 3.

Classification with SVM using the sigmoid nucleus Figures 8 and 9 and Table 4.

According to Figures 2-9 and Tables 1-4, we can say that the use of the RBF nucleus for the phase of learning gives the best results, that is to say, the highest rate of the word classification. The values of the RBF have been adapted to the learning process, because the mathematical function used in the RBF kernel processes the learning values better than the mathematical functions of other kernels (linear, polynomial, and sigmoid).

This is due to the good choice of our database, and the methods used for this classification, which are the MFCCs for speech parameterization, and the SVMs for learning and classification.

For the database, each word is pronounced three times by each speaker; we used 24 speakers: 8 children, 8 men, and 8 women, whose age varies between 17 and 65, for children between 5 years and 16 years, and we repeated these experiences three times, under different conditions, in order to have a robust database.

The classification rate varies between 96.4 and 99.41% and with a precision that varies between 53.11 and 96.96% and a recall that varies between 58.45 and 98.85%.

There are words that have a low precision because

These words are very similar especially in the pronunciation like window on, window off, and in

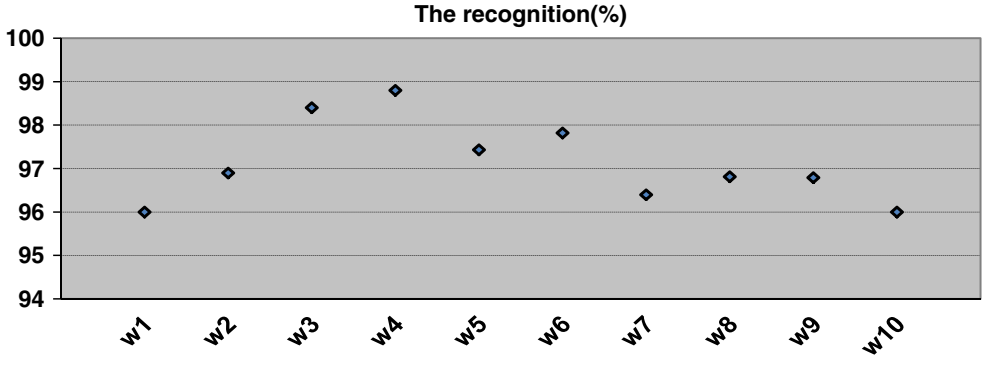


Figure 2: The rate of the recognition of the words w_1 to w_{10} by Mel FCC-SVM using the linear kernel nucleus.

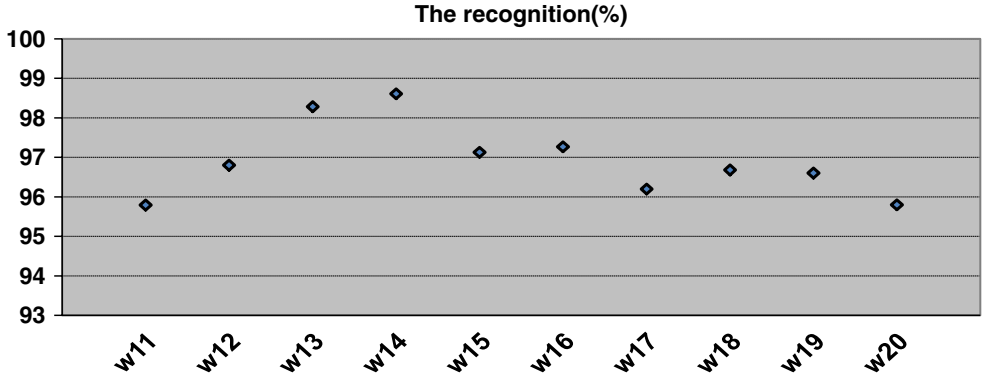


Figure 3: The rate of the recognition of the words w_{11} to w_{20} by Mel FCC-SVM using the linear kernel nucleus.

Table 1. Painting of the recall and precision for the recognition of the words for the linear kernel nucleus.

Word classified	Precision linear kernel (%)	Recall for linear kernel (%)
W1	91.01	92.29
W2	89.12	90.59
W3	94.31	95.89
W4	84.09	86.15
W5	54.96	62.96
W6	50.89	69.90
W7	81.63	85.68
W8	57.96	56.85
W9	78.26	84.23
W10	75.85	80.80
W11	80.19	88.42
W12	69.45	77.94
W13	76.13	82.09
W14	73.48	78.29
W15	78.02	85.96
W16	66.51	60.86
W17	89.63	91.79
W18	87.67	90.69
W19	92.03	95.45
W20	82.04	85.78

this case, the system of the classification of the word, recognizes these words, which decreases the precision of the word to classify.

These words are complicated, like window down, in this case, the speech classification system recognizes it with low precision.

The image classification system

We conducted experiments in the Domus intelligent building at the LIG laboratory, located at the University of Grenoble, France, specifically at the CTL (Software Development Center).

Our objective is to set up a complete video surveillance network that includes two parts: the first part, consists in fixing two cameras in each room of the intelligent building, these are connected to a computer for the acquisition of permanent video sequences. The second part consists in recording the video sequences for 30 seconds, then converting them into images. Figure 10 represents the three images to be classified by the image classification system, I_1 : the presence of the person, I_2: a sleeping person, I_3: the absence of the person.

Database input

The input database which we carried out is carried out starting from eight people (different sizes and sex). We register in the intelligent building the presence of a person, a sleeping person, and the absence of a person (Fig. 10). We conducted again these experiments six times, so that each experiment has a different lighting compared to the others, in order to take into account the variation of the lighting, and to obtain a robust system.

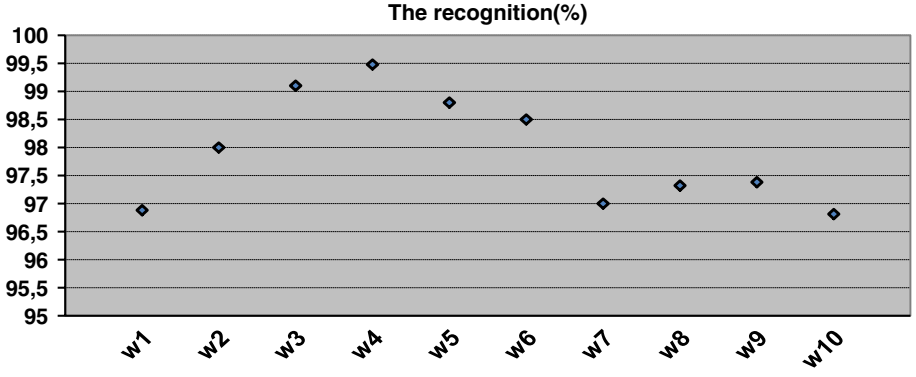


Figure 4: The rate of the recognition of the words w_1 to w_{10} by Mel FCC-SVM using the RBF (radial basic function) nucleus.

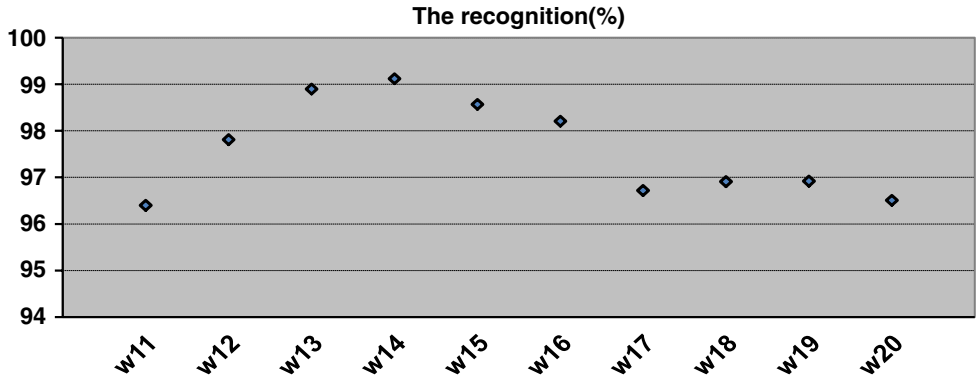


Figure 5: The rate of the recognition of the words w_{11} to w_{20} by Mel FCC-SVM using the RBF (radial basic function) nucleus.

Table 2. Painting of the recall and precision for the recognition of the words for the RBF (radial basic function) nucleus.

Word classified	Precision for RBF kernel (%)	Recall for RBF kernel (%)
W1	92.12	94.02
W2	90.89	92.05
W3	96.96	97.13
W4	86.42	88.85
W5	56.82	64.33
W6	53.11	71.22
W7	83.08	87.13
W8	60.02	58.45
W9	83.26	86.89
W10	80.96	83.58
W11	84.89	90.22
W12	71.08	80.09
W13	78.86	84.92
W14	75.56	80.85
W15	80.22	88.18
W16	68.23	63.08
W17	90.33	93.11
W18	88.45	90.87
W19	94.36	98.85
W20	84.81	88.58

These recordings are made under different lighting. Each video sequence is subdivided into a set of 30-sec video sequences. From each video sequence, we extract a set of representative images.

Each image is made up of 1,152 × 648 pixels. In other words, the corpus that we obtained consists of 1,002 images. In total, 70% of the input corpus is devoted to learning, and the 30% to classification or testing, which gives us 701 images for learning and 300 images for the test.

Parameterization

We then apply to each image the Scale IFT (Scale Invariant Feature Transform) method for the configuration of these images in order to represent each image by a vector of the same size containing 128 values. The implementation takes place in two phases: learning and classification.

$I_{i\,SIFT}$ represents the image I_i by the SIFT method, and I_{ij} the coefficients.

$$I_{i\,SIFT} = \begin{vmatrix} I_{11} & I_{12} & \ldots & I_{1\,16} \\ I_{21} & I_{22} & \ldots & I_{2\,16} \\ I_{31} & I_{32} & \ldots & I_{3\,16} \\ I_{41} & I_{42} & \ldots & I_{4\,16} \\ I_{51} & I_{52} & \ldots & I_{5\,16} \\ I_{61} & I_{62} & \ldots & I_{6\,16} \\ I_{71} & I_{72} & \ldots & I_{7\,16} \\ I_{81} & I_{82} & \ldots & I_{8\,16} \end{vmatrix}$$

Learning the network

For the implementation, we use several kernel functions to take the best models that give us the highest classification rate. Since we have three images to recognize, the SVM will generate three

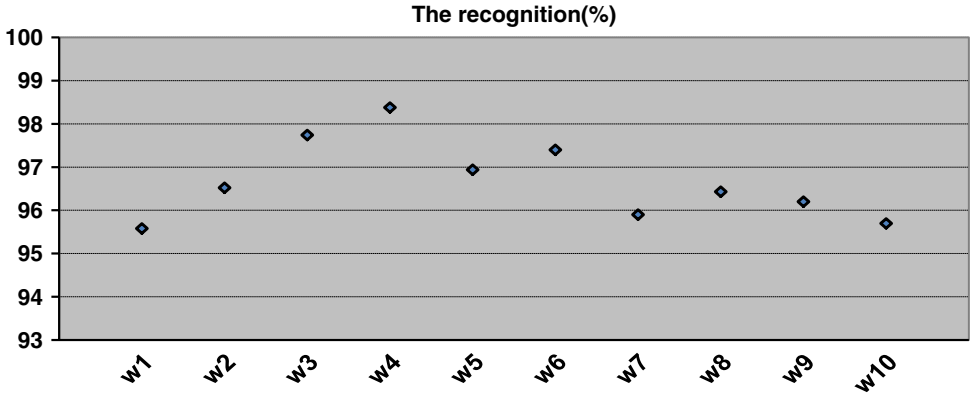


Figure 6: The rate of the recognition of the words w_1 to w_{10} by Mel FCC-SVM using the polynomial kernel nucleus.

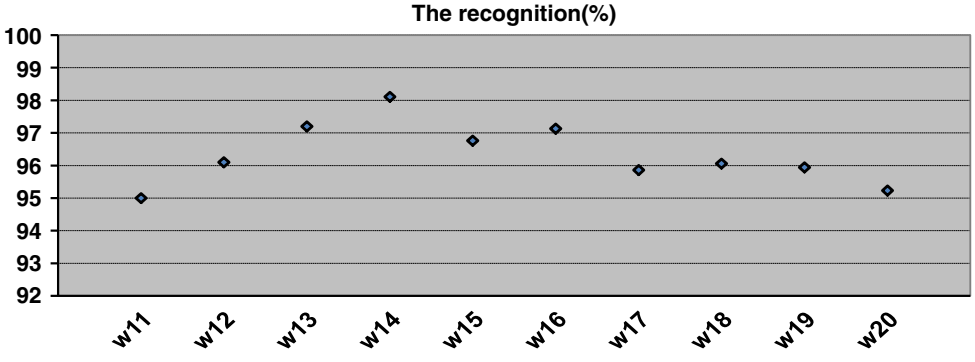


Figure 7: The rate of the recognition of the words w_{11} to w_{20} by Mel FCC-SVM using the polynomial kernel nucleus.

models (classifiers or hyperplane separators) in the learning or training phase, as each model corresponds to one image. In order to improve performance, a study on several nuclei (Linear, RBF (Radial basic function), Polynomial, and Sigmoid) was carried out in the learning phase, taking the nucleus, which gives the best results. By using SVM (vector machine supports), we apply a learning method to reduce errors related to system recognition, followed by a classification technique that determines the three images in the smart building.

Recognition

As we have three images to recognize for the image classification system, then the SVMs will generate three models (classifiers or hyperplane separators) in the learning phase, and each model corresponds to an image. Since we have three classes, then we will have three binary classifiers, which are constructed (three hyperplanes, or three problems of optimization), the third classifier being intended to distinguish the class of index 3. Each image to be classified will be presented to the three classifiers and is assigned the label of the classifier, which returned the highest percentage of adhesion. The diagram of the image classification system by SIFT and SVM is shown in Figure 11.

Resulte discussion

1. The rate of recognition / $I_i = \frac{NRI_i}{N_T I_i}$, (4)

   NRI_i: represents the number of recognized image I_i.
   $N_T I_i$: represents the total number of the image I_i in the base of tets.
   RR: represents the rate of recognition.
2. The precision for a I_i image is the number of images correctly detected in divided by the total

Table 3. Painting of the recall and precision for the recognition of the words for the polynomial kernel nucleus.

Word classified	Precision polynomial kernel (%)	Recall for polynomial kernel (%)
W1	90.20	91.56
W2	88.01	89.74
W3	93.24	94.92
W4	83.11	85.39
W5	53.86	61.99
W6	49.78	68.98
W7	80.33	84.89
W8	56.75	55.93
W9	77.13	83.59
W10	74.71	79.91
W11	79.46	87.75
W12	68.59	76.98
W13	75.22	81.22
W14	72.65	77.36
W15	77.18	84.98
W16	65.76	59.92
W17	88.79	90.84
W18	86.85	89.79
W19	91.26	94.62
W20	81.17	84.81

number of images in the entire test base that are detected and assigned to the image I_i: (https://fr.wikipedia.org/wiki/Pr%C3%A9precision_et_rappel).
$N_{cd}I_i$: the number of images correctly detected for the image I_i.
N_dI_i: the number of images detected for the image I_i in the test base.
So:

$$\text{Precision} / I_i = \frac{N_{cd}I_i}{N_dI_kI_i}. \quad (5)$$

3. The reminder for a I_i image, this is the number of correctly detected images for the I_i image divided by the total number of detected t images that belong to the I_i image (https://fr.wikipedia.org/wiki/Pr%C3%A9precision_et_rappel).
N_{cd} I_i: the number of images correctly detected for the image I_i.
N_T I_i: the total number of detected images that belong to image I_i.
So:

$$\text{Reminder} / I_i = \frac{N_{cd}I_i}{N_T\ I_i}. \quad (6)$$

As long as the values of the recognition rate, precision, and recall are high, we can guarantee its reliability and accuracy.

The classification is done by the SVM. To improve this classification, we used 4 nuclei, linear, polynomial, RBF (radial basic function), and sigmoid.

The results of the experiments are given in Figures 12-15 and Tables 5-8.

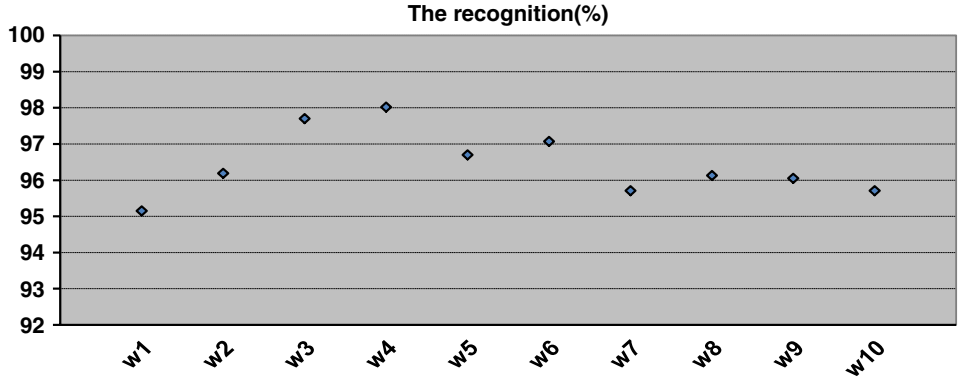


Figure 8: The rate of the recognition of the words w_1 to w_{10} by Mel FCC-SVM using the sigmoid nucleus.

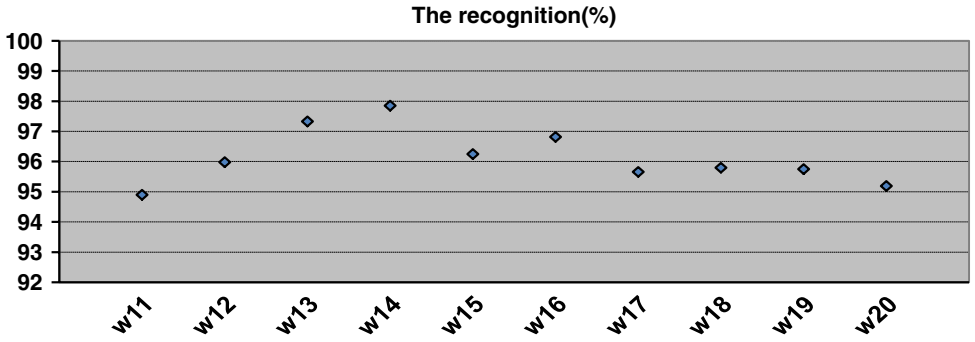
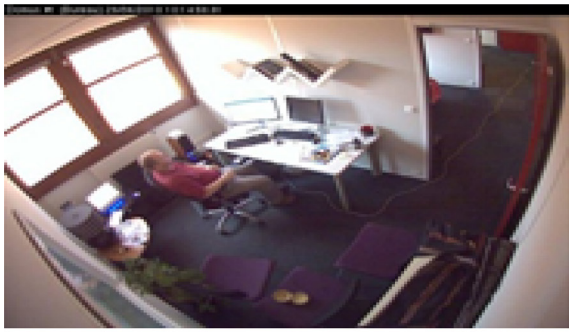


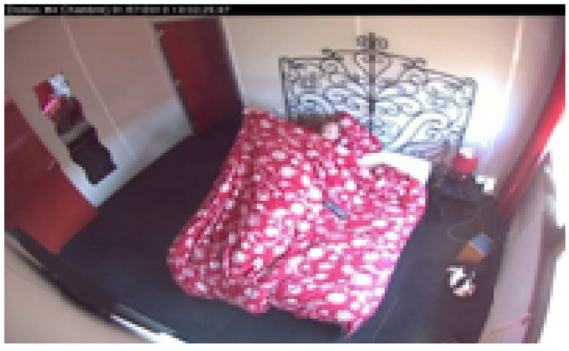
Figure 9: The rate of the recognition of the words w_{11} to w_{20} by Mel FCC-SVM using the sigmoid nucleus.

Table 4. Painting of the recall and precision for the recognition of the words for the sigmoid kernel nucleus.

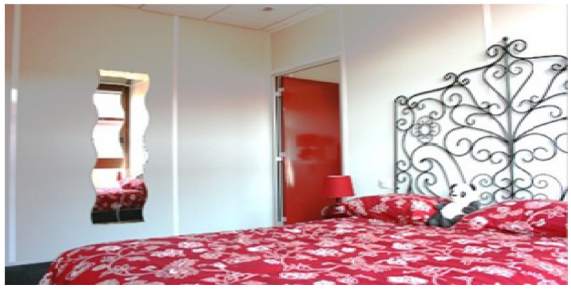
Word classified	Precision for sigmoid kernel (%)	Recall for sigmoid kernel (%)
W1	89.40	90.96
W2	87.25	89.01
W3	92.50	94.11
W4	82.35	84.60
W5	53.01	61.14
W6	49.02	68.12
W7	79.50	84.08
W8	56.02	55.09
W9	76.40	82.79
W10	74.03	79.08
W11	78.65	87.01
W12	67.80	76.11
W13	74.40	80.45
W14	71.80	76.60
W15	76.40	84.15
W16	65.01	59.13
W17	88.02	90.02
W18	86.07	89.01
W19	90.45	93.80
W20	80.49	84.01



I_1



I_2



I_3

Figure 10: Photos of the Domus intelligent building layout, I_1: the person is present, I_2: the person is asleep, I_3: the person is absent.

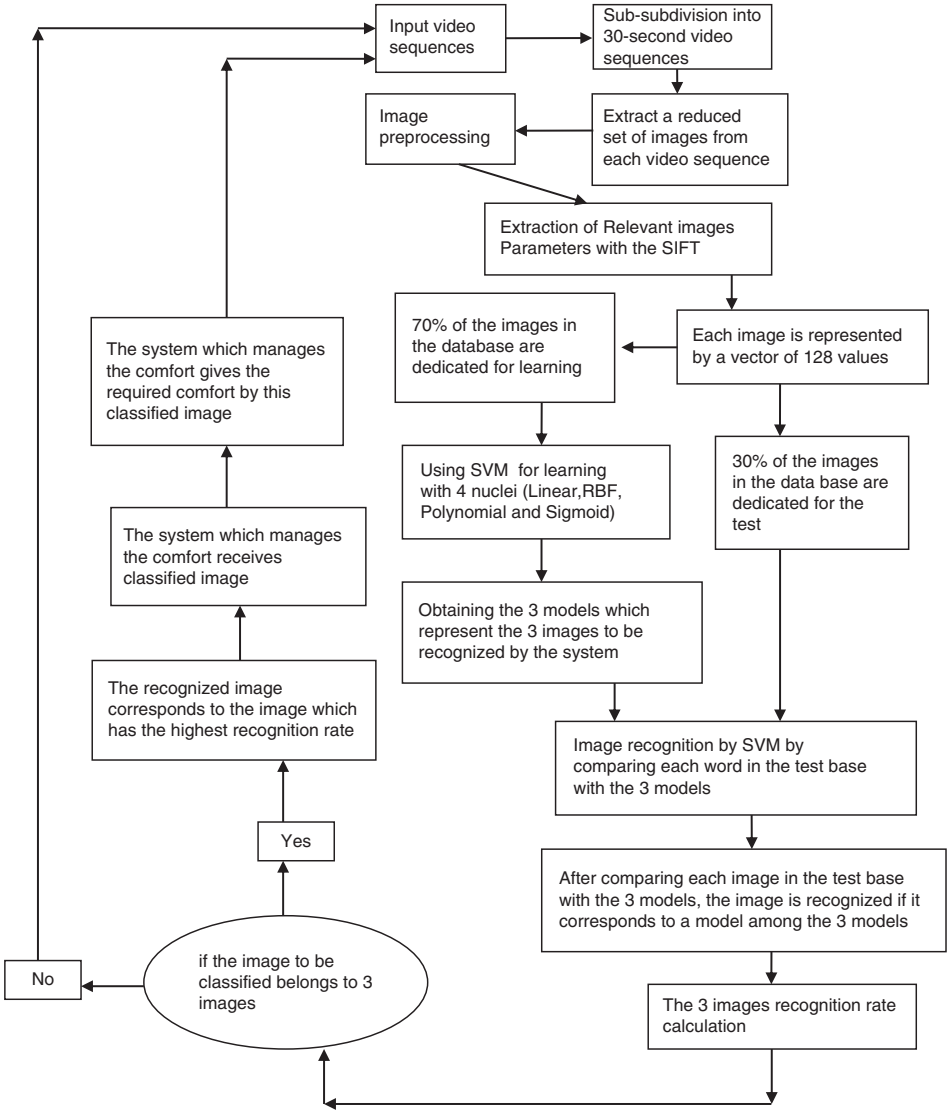


Figure 11: The diagram of the image classification system by SIFT and SVM.

Figure 12: The rate of the recognition of the images by SIFT-SVM using the linear kernel nucleus.

Table 5. Painting of the recall and precision for the recognition of the image for the linear kernel nucleus.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	96.38	96.89
I2	76.42	79.11
I3	93.38	95.16

Note: e.2 Classification with SVM using RBF (radial basic function) nucleus.

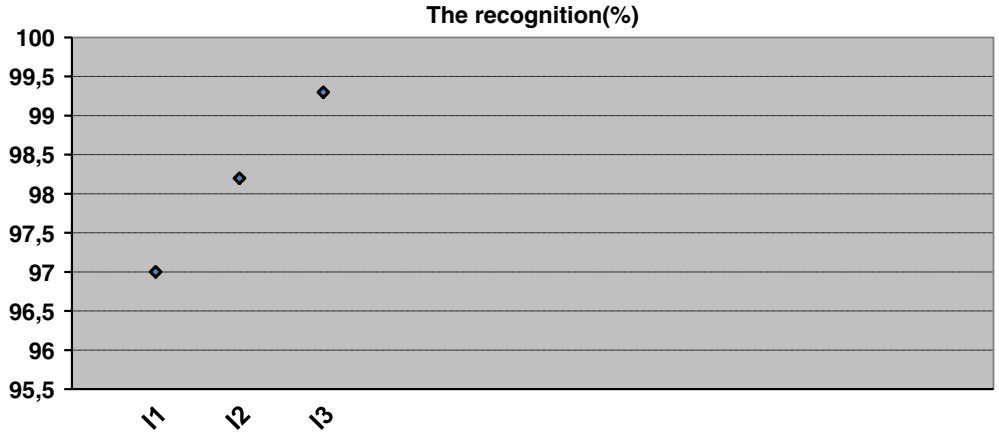


Figure 13: The rate of the recognition of the images by SIFT-SVM using the RBF (radial basic function) nucleus.

Table 6. Painting of the recall and precision for the recognition of the image for the RBF kernel nucleus.

Image classified	Precision RBF kernel (%)	Recall for RBF kernel (%)
I1	97.29	97.96
I2	77.89	80.65
I3	94.45	96.13

Note: e.3 Classification with SVM using the polynomial kernel nucleus.

Classification with SVM using the linear kernel nucleus Figure 12 and Table 5.

Classification with SVM using RBF (radial basic function) nucleus Figure 13 and Table 6.

Classification with SVM using the polynomial kernel nucleus Figure 14 and Table 7.

Classification with SVM using the sigmoid nucleus Figure 15 and Table 8.

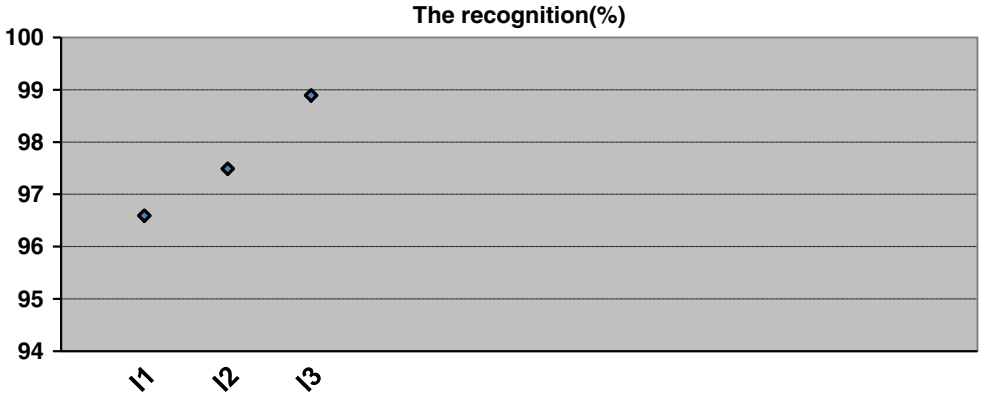


Figure 14: The rate of the recognition of the images by SIFT-SVM using the polynomial kernel nucleus.

Table 7. Painting of the recall and precision for the recognition of the image for the polynomial kernel nucleus.

Image classified	Precision polynomial kernel	Recall for I polynomial kernel
I1	95.56	95.97
I2	75.72	78.56
I3	92.79	94.45

Note: e.4 Classification with SVM using the sigmoid nucleus.

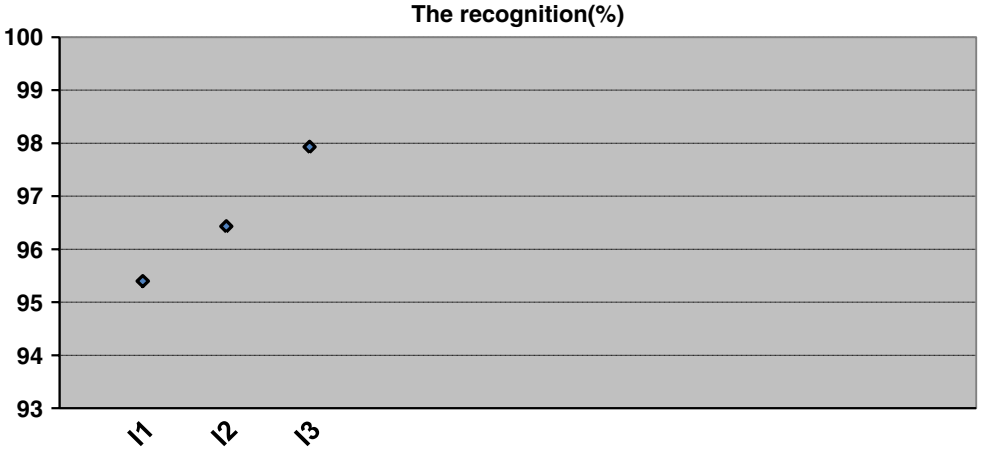


Figure 15: The rate of the recognition of the images by SIFT-SVM using the sigmoid nucleus.

According to Figures 12-15 and Tables 5-8, we can say that the use of the RBF core for the phase of learning gives the best results, it means the highest rate of the Image classification. This is due to the good choice of our database, and the methods used for this classification, which are the Scal IFT for

Table 8. Painting of the recall and precision for the detection of linear behavior for the core.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	94.89	94.99
I2	74.91	77.83
I3	91.94	93.66

image parameterization, and the SVMs for learning and classification. For the database, we register in the intelligent building the presence of a person, the sleeping person, and the absence of a person. These experiments are repeated 6 times, so that each experiment has a different lighting from the others, in order to take into account the variation in lighting. We also used different persons (age, sex), to play the

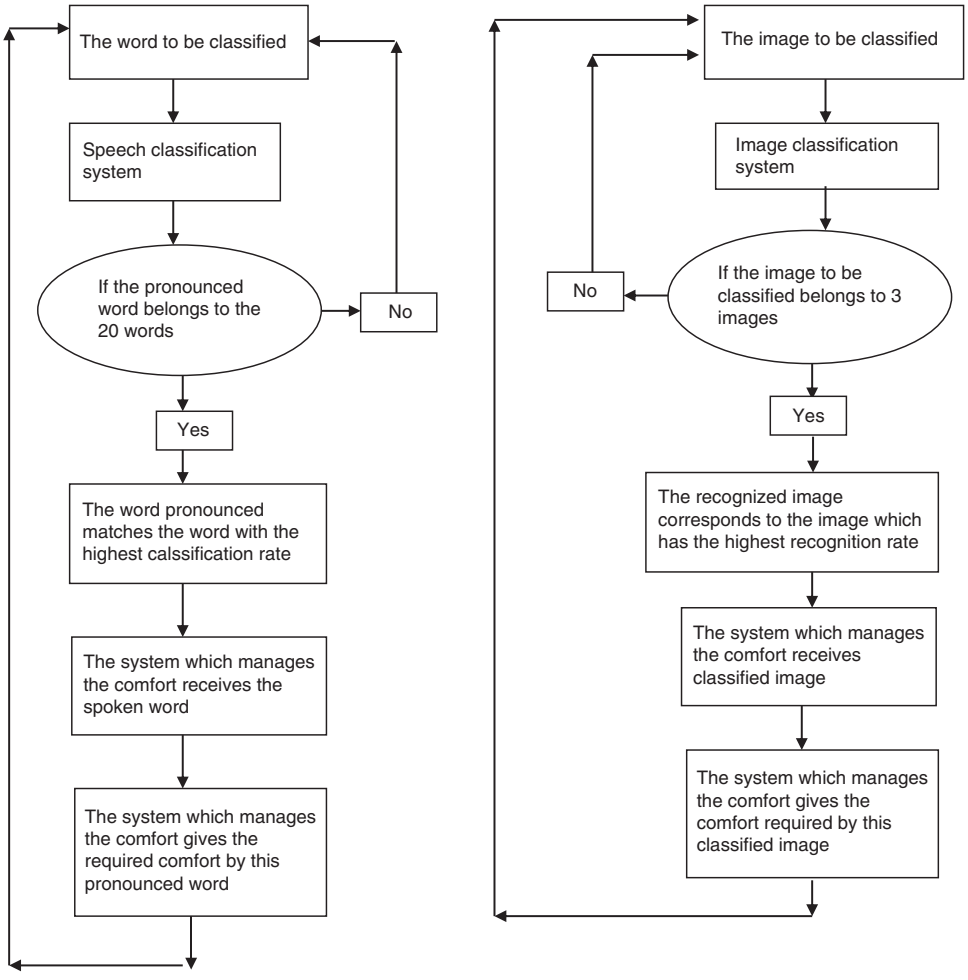


Figure 16: Diagram of the system of the command of comfort in the intelligent building.

scenario that represents a person present, a sleeping person, and an absent person.

The classification rate varies between 97 and 99.3% and with an accuracy that varies between 77.89 and 97.29% and a recall that varies between 80.65 and 97.96%. Images that have lower accuracy and recall have bad lighting, especially all images that express a sleeping person. The images express the presence of a person with the best precision and recall, because in this case we have a very good lighting.

The system of the command of the comfort in the intelligent building

The comfort command system in an intelligent building that we have created, is composed of two parallel systems, which work at the same time.

The first system consists in classifying the word, precisely the classification of 20 words; the second system consists of classifying the images, specifically the classification of three images (Fig. 16). Our system is robust and optimal, because we obtained good results for the precision, the recall, and the rate of the classification, and also, because we used only 20 words for the first system and three images for the second system, which represents an optimal number for the command of comfort. The good results obtained for the recognition rate, the precision, and the recall for both systems, are the condition to be taken into account for reliability and accuracy in this methodology.

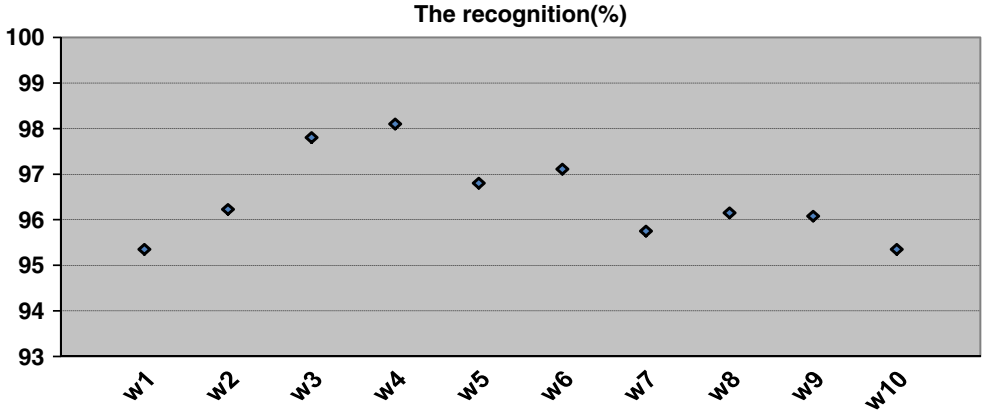


Figure 17: The rate of the recognition of the words w_1 to w_{10} by Linear PC-SVM using the linear kernel nucleus.

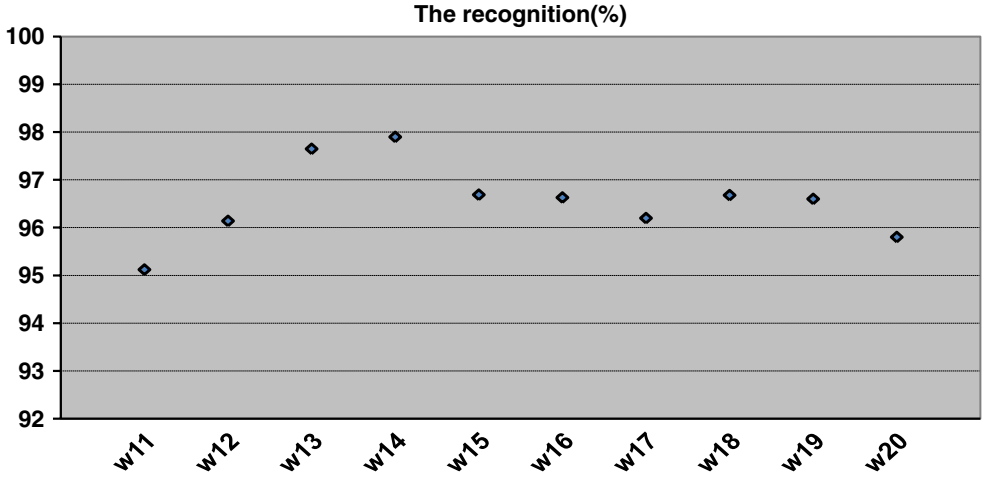


Figure 18: The rate of the recognition of the words w_{11} to w_{20} by LPC-SVM using the linear kernel nucleus.

Table 9. Painting of the recall and precision for the recognition of the words for the linear kernel nucleus.

Word classified	Precision linear kernel	Recall for linear kernel
W1	90.36	91.75
W2	88.63	89.93
W3	93.75	95.13
W4	83.47	85.59
W5	54.23	62.19
W6	50.17	69.21
W7	81.01	85.01
W8	57.21	56.17
W9	77.69	83.65
W10	75.11	80.09
W11	79.63	87.67
W12	68.82	77.18
W13	75.56	81.45
W14	72.86	77.63
W15	77.32	85.17
W16	65.83	60.14
W17	88.93	91.04
W18	87.01	90.01
W19	91.72	94.72
W20	81.31	85.03

Comparative analysis

For the speech classification system

In order to validate our obtained results, a comparative study is essential, in order to better see the contribution of the MFCCs of the SVMs for the classification of speech. We repeated the same experiments with the same conditions, but we changed the method of the parameterization of words, such as we used the Linear PC (linear prediction coding) method. In this method, each word will be divided into 21 segments. Each segment will undergo a 23-ms HAMMING windowing (256 samples) with a 40% overlap. From each part 12 Linear PC coefficients are extracted. Order 12 is sufficient to characterize a part. After concatenation of the different blocks of results, each word will be represented by 252 coefficients (21 segments coded with 13 coefficients). $W_{i\ Mel\ FCC}$ represents the word w by the Mel FCC method, and $W_{i\,j}$ the coefficients.

$$W_{i\ Linearl\ PC} = \begin{matrix} W_{1\,1}\ W_{1\,2} \ldots\ldots W_{1\,21} \\ W_{2\,1}\ W_{2\,2} \ldots\ldots W_{2\,21} \\ W_{3\,1}\ W_{3\,2} \ldots\ldots W_{3\,21} \\ W_{4\,1}\ W_{4\,2} \ldots\ldots W_{4\,21} \\ W_{5\,1}\ W_{5\,2} \ldots\ldots W_{5\,21} \\ W_{6\,1}\ W_{6\,2} \ldots\ldots W_{6\,21} \\ W_{7\,1}\ W_{7\,2} \ldots\ldots W_{7\,21} \\ W_{8\,1}\ W_{8\,2} \ldots\ldots W_{8\,21} \\ W_{9\,1}\ W_{9\,2} \ldots\ldots W_{9\,21} \\ W_{10\,1}\ W_{10\,2} \ldots\ldots W_{10\,21} \\ W_{11\,1}\ W_{11\,2} \ldots\ldots W_{11\,21} \\ W_{12\,1}\ W_{12\,2} \ldots\ldots W_{12\,21} \end{matrix}$$

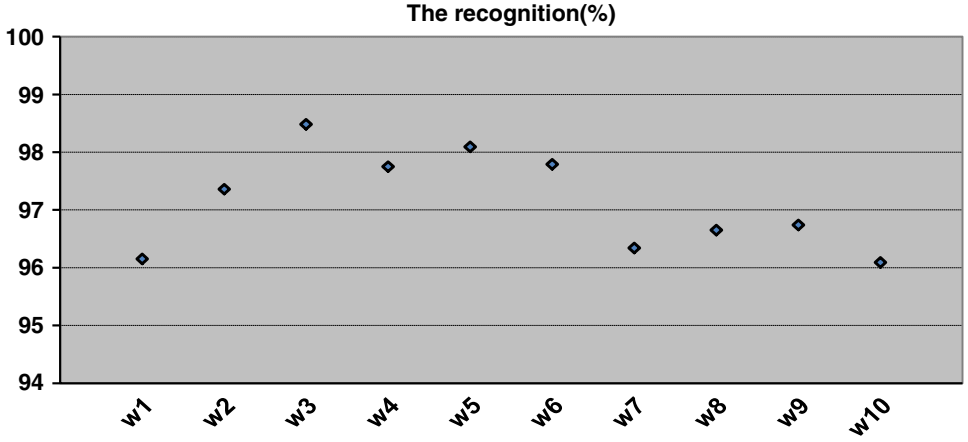


Figure 19: The rate of the recognition of the words w_1 to w_{10} by LPC-SVM using the RBF (radial basic function) nucleus.

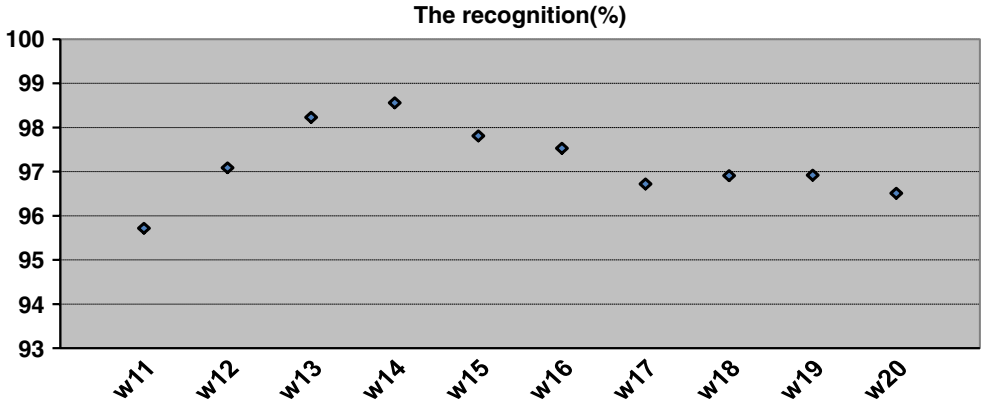


Figure 20: The rate of the recognition of the words w_{11} to w_{20} by LPC-SVM using the RBF (radial basic function) nucleus.

Table 10. Painting of the recall and precision for the recognition of the words for the RBF (radial basic function) nucleus.

Word classified	Precision for RBF kernel (%)	Recall for RBF kernel (%)
W1	91.59	93.31
W2	90.17	91.39
W3	96.21	96.59
W4	85.76	88.13
W5	56.11	63.69
W6	52.56	70.51
W7	82.45	86.59
W8	59.32	57.72
W9	82.51	86.17
W10	80.23	82.89
W11	84.17	89.52
W12	70.46	79.31
W13	78.14	84.21
W14	74.86	80.14
W15	79.53	87.59
W16	67.54	62.45
W17	89.69	92.45
W18	87.79	90.13
W19	93.72	98.14
W20	84.09	87.75

We present the results obtained by the Linear pc method and SVMs for the classification of speech in Figures 17-24 and Tables 9-12.

Classification with SVM using the linear kernel nucleus Figures 17 and 18 and Table 9.

Classification with SVM using RBF (radial basic function) nucleus Figures 19 and 20 and Table 10.

Classification with SVM using the polynomial kernel nucleus Figures 21 and 22 and Table 11.

Classification with SVM using the sigmoid nucleus Figures 23 and 24 and Table 12.

From the results obtained in Figures 17-24 and Tables 9-12, we can say that our results obtained are better than the results obtained with the Linear PC method and the SVMs.

For the image classification system

A comparative study is essential to validate our obtained results for the image classification that uses Scal IFT and SVM methods. We conducted again the same experiments with the same conditions, but we changed the method of the parameterization of images, and we used two different methods, which are Local BP (Local Binary Pattern) histograms and Red GB (red green blue) color histograms.

Local BP histograms

We then apply to each image the Local BP (Local Binary Pattern) method for the configuration of these images in order to represent each image by a vector of the same size containing 64 values. The implementation takes place in two phases: learning and classification.

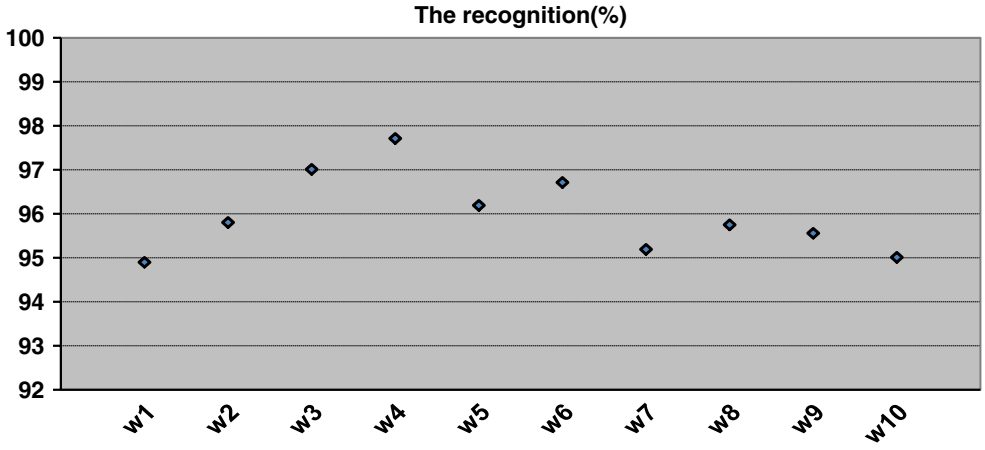


Figure 21: The rate of the recognition of the words w_1 to w_{10} by LPC-SVM using the polynomial kernel nucleus.

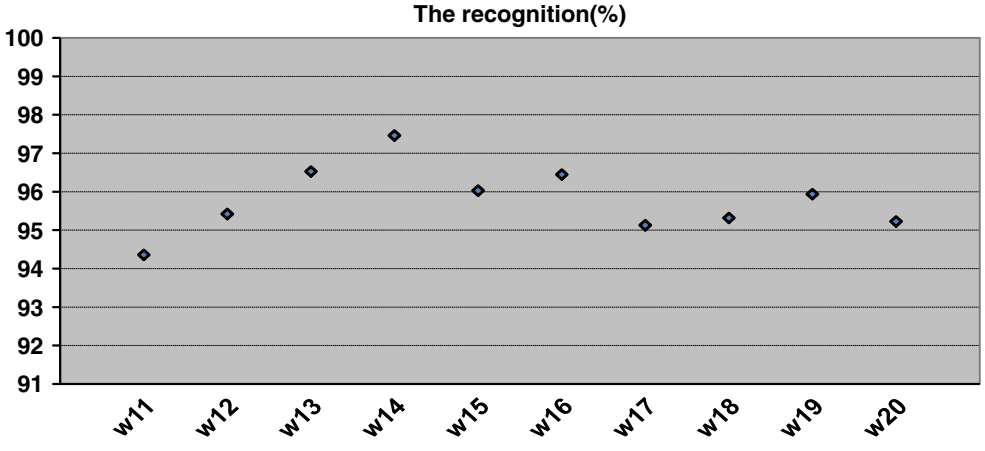


Figure 22: The rate of the recognition of the words w_{11} to w_{20} by LPC-SVM using the polynomial kernel nucleus.

$I_{i\ Local\ PC}$ represents the image I_i by the Local BP method, and I_{ij} the coefficients.

$$I_{i\ Local}\ BP = \begin{matrix} I_{11} & I_{12} & \ldots & I_{18} \\ I_{21} & I_{22} & \ldots & I_{28} \\ I_{31} & I_{32} & \ldots & I_{38} \\ I_{41} & I_{42} & \ldots & I_{48} \\ I_{51} & I_{52} & \ldots & I_{58} \\ I_{61} & I_{62} & \ldots & I_{68} \\ I_{71} & I_{72} & \ldots & I_{78} \\ I_{81} & I_{82} & \ldots & I_{88} \end{matrix}$$

We present the results obtained by the Local BP method and SVMs for the classification of image in Figures 25-28 and Tables 13-16.

Classification with SVM using the linear kernel nucleus Figure 25 and Table 13.

Classification with SVM using RBF (radial basic function) nucleus Figure 26 and Table 14.

Classification with SVM using the polynomial kernel nucleus Figure 27 and Table 15.

Classification with SVM using the sigmoid nucleus Figure 28 and Table 16.

From the results obtained in Figures 25-28 and Tables 13-16, we can say that our results obtained are better than the results obtained with the Local BP method and the SVMs.

Red GB histograms

We then apply to each image the Red GB (Red green blue) method for the configuration of these images in order to represent each image by a vector of the same size containing 256 values. The

Table 11. Painting of the recall and precision for the recognition of the words for the polynomial kernel nucleus.

Word classified	Precision polynomial kernel (%)	Recall for polynomial kernel (%)
W1	89.51	90.84
W2	87.32	89.01
W3	92.50	94.15
W4	83.11	84.71
W5	53.11	61.21
W6	49.06	68.25
W7	79.62	84.14
W8	56.01	55.19
W9	76.48	82.91
W10	74.02	79.18
W11	78.70	87.02
W12	67.92	76.21
W13	74.52	80.56
W14	71.93	76.63
W15	76.49	84.26
W16	65.03	59.19
W17	88.01	90.12
W18	86.12	89.03
W19	90.56	93.89
W20	80.48	84.08

implementation takes place in two phases: learning and classification.

$I_{i\ Red\ GB}$ represents the image I_i by the Red GB method, and I_{ij} the coefficients.

$$I_{i\ Red\ GB} = \begin{vmatrix} I_{1\,1} & I_{1\,2} & \cdots & I_{1\,16} \\ I_{2\,1} & I_{2\,2} & \cdots & I_{2\,16} \\ I_{3\,1} & I_{3\,2} & \cdots & I_{3\,16} \\ I_{4\,1} & I_{4\,2} & \cdots & I_{4\,16} \\ I_{5\,1} & I_{5\,2} & \cdots & I_{5\,16} \\ I_{6\,1} & I_{6\,2} & \cdots & I_{6\,16} \\ I_{7\,1} & I_{7\,2} & \cdots & I_{7\,16} \\ I_{8\,1} & I_{8\,2} & \cdots & I_{8\,16} \\ I_{9\,1} & I_{9\,2} & \cdots & I_{9\,16} \\ I_{10\,1} & I_{10\,2} & \cdots & I_{10\,16} \\ I_{11\,1} & I_{11\,2} & \cdots & I_{11\,16} \\ I_{12\,1} & I_{12\,2} & \cdots & I_{12\,16} \\ I_{13\,1} & I_{13\,2} & \cdots & I_{13\,16} \\ I_{14\,1} & I_{14\,2} & \cdots & I_{14\,16} \\ I_{15\,1} & I_{15\,2} & \cdots & I_{15\,16} \\ I_{16\,1} & I_{16\,2} & \cdots & I_{16\,16} \end{vmatrix}$$

We present the results obtained by the Red GB method and SVMs for the classification of image in Figures 29-32 and Tables 17-20.

Classification with SVM using the linear kernel nucleus Figure 29 and Table 17.

Classification with SVM using RBF (radial basic function) nucleus Figure 30 and Table 18.

Classification with SVM using the polynomial kernel nucleus Figure 31 and Table 19.

Classification with SVM using the sigmoid nucleus Figure 32 and Table 20.

From the results obtained in Figures 29-32 and Tables 17-20, we can say that our results obtained

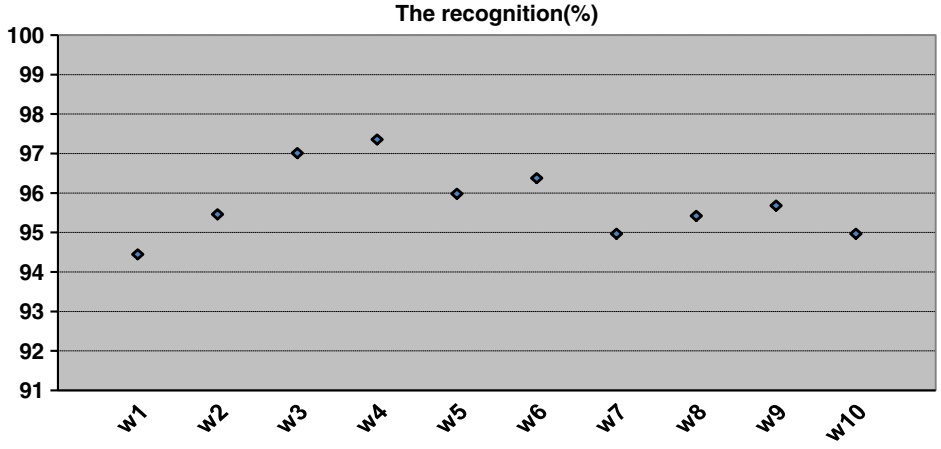


Figure 23: The rate of the recognition of the words w_1 to w_{10} by LPC-SVM using the sigmoid nucleus.

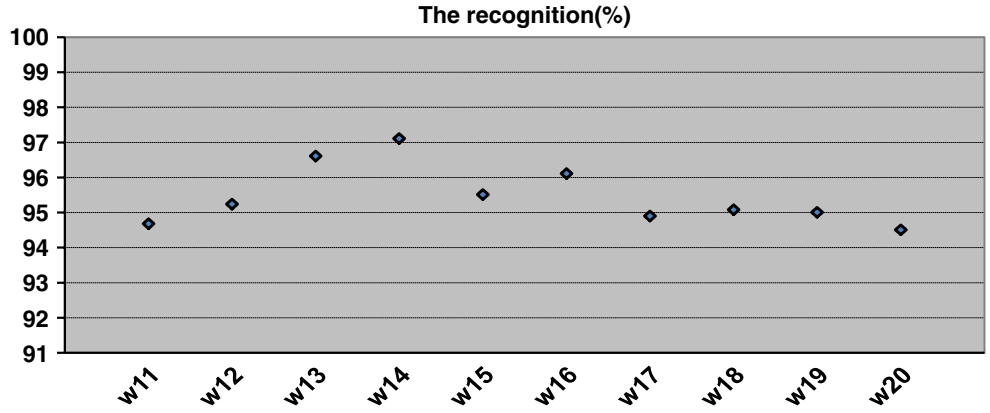


Figure 24: The rate of the recognition of the words w_{11} to w_{20} by LPC-SVM using the sigmoid nucleus.

Table 12. Painting of the recall and precision for the recognition of the words for the sigmoid kernel nucleus.

Word classified	Precision for sigmoid kernel (%)	Recall for sigmoid kernel (%)
W1	88.70	90.21
W2	86.51	88.27
W3	91.75	93.39
W4	81.60	83.90
W5	52.36	60.39
W6	48.32	67.40
W7	78.83	83.31
W8	55.32	54.30
W9	75.72	82.01
W10	73.34	78.35
W11	77.90	86.28
W12	67.08	75.39
W13	73.72	79.69
W14	71.07	75.86
W15	75.75	83.39
W16	64.32	58.40
W17	87.28	89.27
W18	85.29	88.25
W19	89.72	93.07
W20	79.75	83.24

are better than the results obtained with the Red GB method and the SVMs.

Conclusion

Comfort control is seen as a solution to optimize energy in the intelligent building. There are several methods to control this comfort, among these methods, we have the classification of speech, which gives the person who occupies the intelligent building the possibility of obtaining the desired comfort using his voice, and in this way the comfort will be optimized. But the disadvantages of speech classification are (1) if the person is sick and cannot speak, then in this case the speech classification system will not be able to command comfort. Energy will not be optimized. (2) If the person sleeps leaving the lights on, the television, or music on, then in this case, energy will not be optimized, because these equipments will not be off as long as that person is asleep. In order to overcome these drawbacks, we proposed to add another system of image classification in parallel to the system of speech classification.

This image classification system will be used to classify three images that express the presence of the person, a sleeping person, and the absence of the person. When the person is sick and cannot speak, then in this case, the image classification system will detect the presence of the person, and the intelligent building comfort management system will give an optimal comfort to the person. This person must be satisfied with this optimal comfort, because the speech classification system will not recognize his voice.

In the event that the image classification system does not detect the presence, the building comfort

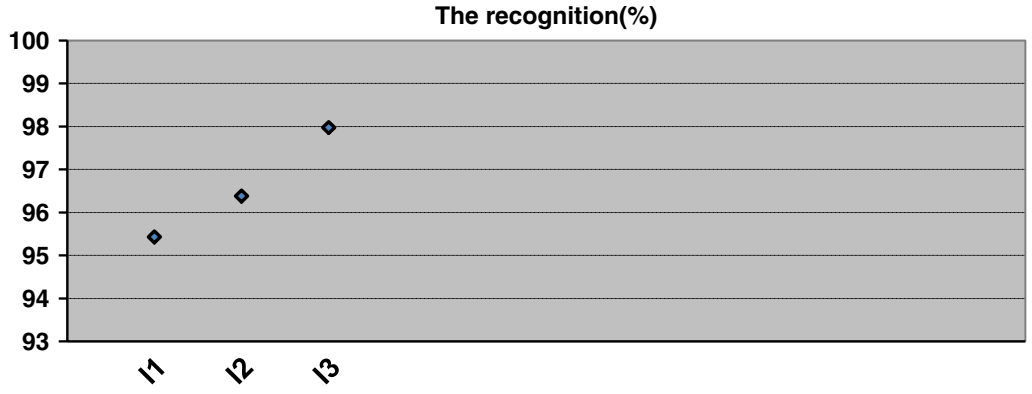


Figure 25: The rate of the recognition of the images by LBP-SVM using the linear kernel nucleus.

Table 13. Painting of the recall and precision for the recognition of the image for the linear kernel nucleus.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	95.90	96.41
I2	75.91	78.60
I3	92.80	94.62

Note: e.2 Classification with SVM using RBF (radial basic function) nucleus.

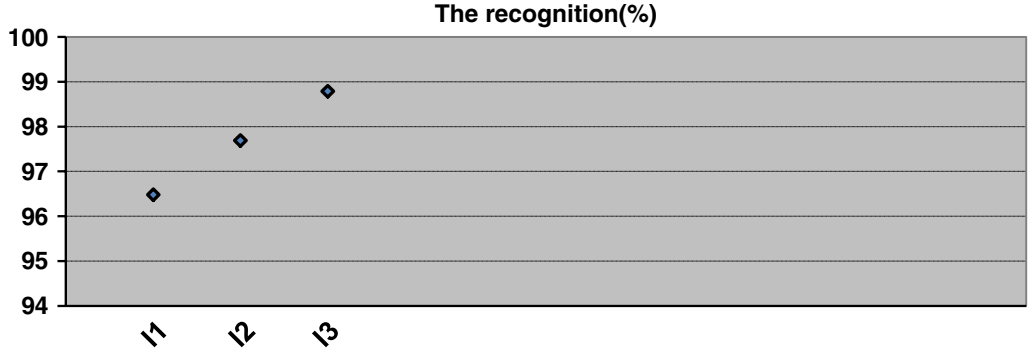


Figure 26: The rate of the recognition of the images by LBP-SVM using the RBF (radial basic function) nucleus.

management system turns off all lights and other equipment in that room, and lowers the heating or air conditioning.

In the event that the image classification system detects that the person is asleep, then the intelligent building comfort management system turns off all lights and other equipment in that room.

It can be said that the addition of this image classification system has solved the drawbacks of the automatic speech recognition system. But the image

Table 14. Painting of the recall and precision for the recognition of the image for the RBF kernel nucleus.

Image classified	Precision RBF kernel (%)	Recall for RBF kernel (%)
I1	96.70	97.41
I2	77.32	80.08
I3	93.85	95.58

Note: e.3 Classification with SVM using the polynomial kernel nucleus.

Figure 27: The rate of the recognition of the images by LBP-SVM using the polynomial kernel nucleus.

Table 15. Painting of the recall and precision for the recognition of the image for the polynomial kernel nucleus.

Image classified	Precision polynomial kernel (%)	Recall for I polynomial kernel (%)
I1	95.01	95.38
I2	75.14	78.02
I3	92.19	93.88

Note: e.4 Classification with SVM using the sigmoid nucleus.

classification system has some drawbacks, such as bad lighting, darkness, and shadow problem. The automatic voice recognition system will solve these drawbacks, such as allowing the occupant of the smart building to stay in the dark, or have poor lighting while still being comfortable.

In the experiments carried out, we have used for the two systems, 4 cores (linear, RBF, polynomial, and

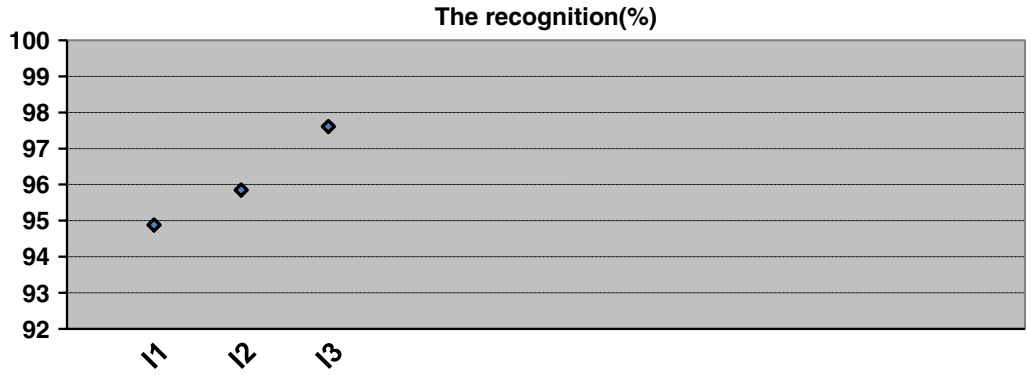


Figure 28: The rate of the recognition of the images by LBP-SVM using the sigmoid nucleus.

Table 16. Painting of the recall and precision for the detection of linear behavior for the core.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	94.37	94.41
I2	74.39	77.28
I3	91.40	93.04

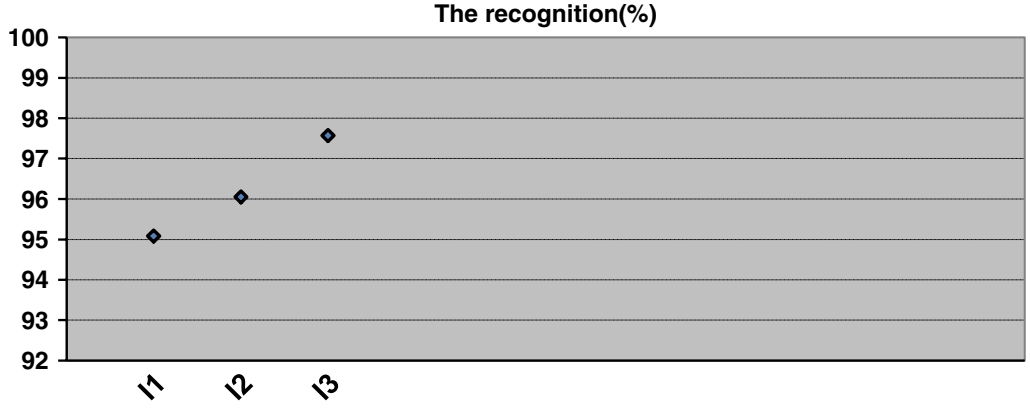


Figure 29: The rate of the recognition of the images by RGB-SVM using the linear kernel nucleus.

sigmoid) in order to choose the best model that gives the best results for the classification. The speech and the image classification system have given satisfactory results especially for the RBF kernel as for the speech classification system the classification rate varies between 96.4 and 99.41% and with a precision that varies between 53.11 and 96.96% and a recall that varies between 58.45 and 98.85%. For the image classification system, the classification rate varies between 97 and 99.3% and with an accuracy

Table 17. Painting of the recall and precision for the recognition of the image for the linear kernel nucleus.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	95.42	95.96
I2	75.49	78.29
I3	92.47	94.28

Note: e.2 Classification with SVM using RBF (radial basic function) nucleus.

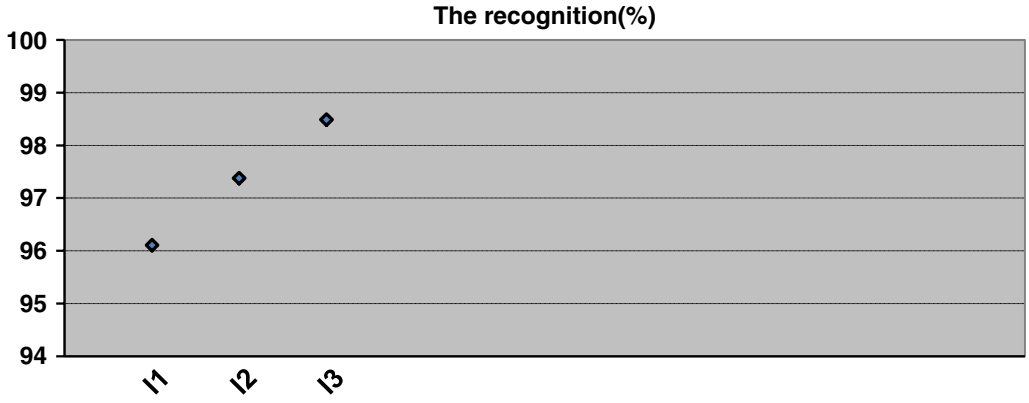


Figure 30: The rate of the recognition of the images by RGB-SVM using the RBF (radial basic function) nucleus.

Table 18. Painting of the recall and precision for the recognition of the image for the RBF kernel nucleus.

Image classified	Precision RBF kernel	Recall for RBF kernel
I1	96.40	96.99
I2	76.98	79.89
I3	93.59	95.24

Note: e.3 Classification with SVM using the polynomial kernel nucleus.

that varies between 77.89 and 97.29% and a recall that varies between 80.65 and 97.96%.

In conclusion, we can say that our goal has been achieved because we have realized a robust and efficient system for the control of comfort in the intelligent building.

From the comparative study that we carried out, we can say that the experimental results obtained are satisfactory for the methods used, which allows us to say that Mel FCC and SVM are recommended for the classification of speech, and Scal IFT and SVM are recommended for image classification,

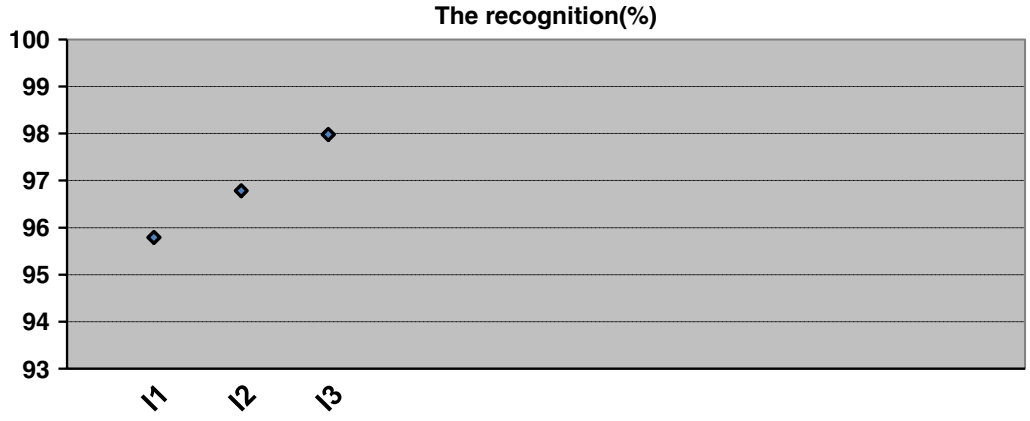


Figure 31: The rate of the recognition of the images by RGB-SVM using the polynomial kernel nucleus.

Table 19. Painting of the recall and precision for the recognition of the image for the polynomial kernel nucleus.

Image classified	Precision polynomial kernel (%)	Recall for I polynomial kernel (%)
I1	94.78	95.01
I2	74.84	77.71
I3	91.90	93.69

Note: e.4 Classification with SVM using the sigmoid nucleus.

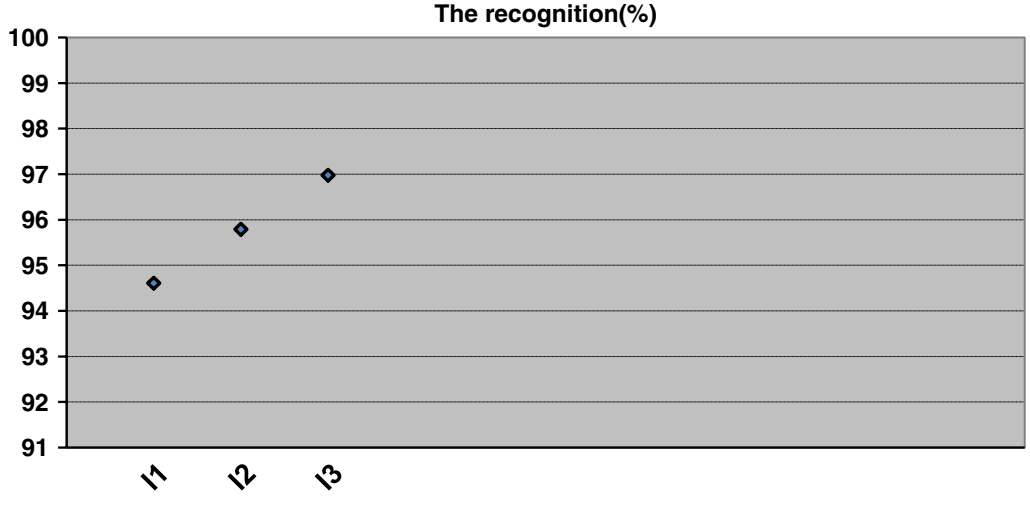


Figure 32: The rate of the recognition of the images by RGB-SVM using the sigmoid nucleus.

Table 20. Painting of the recall and precision for the recognition of the image for the sigmoid kernel nucleus.

Image classified	Precision linear kernel (%)	Recall for linear kernel (%)
I1	93.97	94.02
I2	73.05	77.05
I3	91.11	92.89

and therefore these methods are recommended for comfort command in an intelligent building.

Literature Cited

Fanger, P. O. 2009. Assessment of thermal comfort practice. *Occupational and Environmental Medicine* 313–324.

Givoni, B. and Izard, L. J. 1978. L, l'architecture et le climat. *Éditions du Moniteur* 20–39.

Saizmaa, T. and Kim, H. C. 2008. Smart home design: home or house. Proceedings of the Third International Conference on Convergence an Hybrid Information Technology, Vol. 1, IEEE Computer Society, Washington, DC, pp. 143–148.

Ekambi Schmidt, J. 1972. La perception de l'habitat, Editions Universitaires, p. 186.

Moser, G. 2009. Psychologie environnementale: Les relations homme-environnement De Boeck Université, pp. 272–273.

Cavazza, M., Camara, R. S. and Turunen, M. 2010. How was your day?: a companion eca, in AAMAS, pp. 1629–1630.

Rougui, J., Istrate, D. and Souidene, W. 2009. Audio sound event identification for distress situations and context awareness. EMBC Annual International Conference of the IEEE, Minneapolis, pp. 3501–3504.

Gemmeke, J. F., Ons, B., Tessema, N., Van Hamme, H., Van De Loo, J., De Pauw, G., Daelemans Huyghe, W. J., Derboven, J., Vuegen, L., VanDenBroeck, B., Karsmakers, P. and Vanrumste, B. 2013. Self-taught assistive vocal interfaces: an overview of the aladin project. Inter speech, pp. 2039–2043.

Casanueva, I., Christensen, H., Hain, T. and Green, P. 2014. Adaptive speech recognition and dialogue management for users with speech disorders. Proceedings of Inter speech, Singapore, pp. 1033–1037.

Hamill, M., Young, V., Boger, J. and Mihailidis, A. 2009. Development of an automated speech recognition interface for personal emergency response systems. *Journal of Neuro Engineering and Rehabilitation* 6(1): 1–26.

Ravanelli, M. and Omologo, M. 2014. On the selection of the impulse responses for distant-speech recognition based on contaminated speech training. Proceedings of Inter speech pages, Singapore, pp. 1028–1032.

Ahmed, H. S., Faouzi, B. M. and Caelen, J. 2013. Detection and classification of the behavior ofpeople in an intelligent building by camera. *International Journal on Smart Sensing and Intelligent Systems* 6(4): 1317–1342.

Joachims, T. 1999. "Making large-scale SVM learning practical", In Scholkopf, B., Burges, C. J. C and Smola, A. J. (Eds), *Advances in Kernel Methods – Support Vector Learning* MIT Press, Cambridge, MA, 169–184.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory* Springer, Verla, New York.

Srinivasan, S. and Rajakumar, K. 2017. A review on multiple feature based adaptive sparse representation (MFASR) and other classification types. *International Journal on Smart Sensing and Intelligent Systems* 10(3): 568–571.

Dixon, B. and Candade, N. 2008. Multispectral land use classification using neural networks and support vector machines: one or the other, or both. *International Journal of Remote Sensing* 29(4): 1185–1206.

Karush, W. 1939. Minima of functions of several variables with inequalities as side constraint. Master's thesis, Dept. of Mathematics, Univ. of Chicago, pp. 3–25.

Prasad, S. V. S., Satya Savithri, T., Iyyanki Murali, V. and Krishna, S. 2017. Performance evaluation of SVM kernels on multispectral LISS III data for object classification. *International Journal on Smart Sensing and Intelligent Systems* 10(4): 864.