# New Learning Approach by Co-training for Complex Data Classification

Salma Kammoun Jarraya[1], Rania Rebai Boukhriss[2], Emna Jammouci Fendri[3], Mohamed Hammami[3]
Hanêne Ben-Abdallah[4]
[1] MIRACL Laboratory, Monastir University, ENIM, Av. Ibn EL Jazzar, 5019
Monastir, Tunisia
[2] MIRACL Laboratory, Sfax University, ISIMS, Technologic pole of Sfax: Rte Tunis Km 10 B.P. 242 Sfax 3021
Tunisia
[3] MIRACL Laboratory, Sfax University, FSS
Rte Soukra Km 3 BP 802
3018 Sfax, Tunisia
[4] MIRACL Laboratory, Sfax University
FSEGS, Rte Aeroport Km 4
3018 Sfax, Tunisia
{Salma.Kammoun,Hanene.BenAbdallah}@fsegs.rnu.tn, rania.rebai@hotmail.fr, fendri.msf@gnet.tn
Mohamed.Hammami@fss.rnu.tn

***ABSTRACT:*** *The complexity, abundance and diversity of data (text, images, sound and video) are at the origin of new software solutions for their exploitation. These solutions make use of learning techniques. Among of these techniques, we distinguish learning through co-training, which turned out to be a very exploited by researchers thanks to its low sensitivity to noise, robustness and rapidity. However, the classical variant of co-training showed deficiencies in the classification of complex data particularly in terms of adaptability. In this paper, we present a new approach of semi-supervised learning by Co-training technique to generate a generic classifier from a representative database (MGLP) which is automatically labeled online. Robustness and rapidity, of our approach, to classify complex data were demonstrated through a comparison between the classical variant and the new one in terms of rate and time.*

**Keywords:** Data Mining, Semi-supervised Learning, Co-training

**Received:** 11 November 2012, Revised 22 December 2012, Accepted 30 December 2012



## 1. Introduction

Nowadays, the digital data take a growing place in everyday life and in the professional world. Seen the complexity of their contents, it became essential to propose software solution to extract strategic knowledge. Learning techniques have emerged to extract the useful knowledge for purposes of prediction and decision-making. These techniques can be classified into three main categories: supervised learning, unsupervised learning and semi-supervised learning. Supervised learning means that data can be labeled, or more explicitly the number of classes can be defined. The classification accuracy depends mainly on the size of training data set. The major drawback is that a huge manual labeling database can be tedious. On the other hand, in the case of unsupervised learning, the number of classes is not defined; it could be able to split data in different cluster. The Semi-Supervised Learning (SSL) is mixture a supervised learning which is based on the labeled data and unsupervised learning which uses only unlabeled data. Indeed, the idea of SSL is to use a small set of labeled data and a large set of unlabeled data to

generate a classifier from an automatically labeled training set. Therefore, Semi Supervised Learning becomes a practical interest in several areas.

Moreover, according to the work of [1] [2] [3], there are two reasons for using the SSL: (i) it prevents the labeling of a large data set and (ii) improve prediction accuracy by the use of unlabeled data in learning.

The second reason agrees closely with the classification of complex data. Complex data are those that have similar characteristics between classes or those with characteristics that vary frequently. For example, in case of data from a video stream, the related variety of scenes and changes (abrupt or gradual illumination changes, acquisition noise, etc..) cannot be full viewed in a supervised learning, hence learning must be adaptive to the scene observed. SSL allows "*a priori*" to address the problem of lack of adaptability discovered in supervised learning. Several semi-supervised learning techniques have been proposed in the literature, among these techniques, the Co-training has showed to be the most used. In order to have a suitable technique for complex data, we propose in this paper a new approach to semi-supervised learning by co-training.

The remainder of this paper is organized into four sections. In Section 2, we review the techniques of semi-supervised learning. In Section 3, we focus on the Co-training: we present both classical variant and the proposed one. The purpose and applicability of the proposed approach are validated through an experimental study presented in Section 4. Finally, we end this paper with a conclusion and reflections on the possible uses of the proposed approach.

**2. Semi-supervised learning techniques**

In the literature we distinguish different semi-supervised techniques; the most recognized are self-learning, co-training, semi-supervised support vector machine (S3VM) and generative models. A comprehensive and thorough study of these techniques can be found in [4].

• Self-learning is to train a classifier with labeled data [4]. The classifier is then used to label the data. Most confidantes' labeled data are added to the training dataset. The classifier is re-trained on the labeled data. The procedure is repeated until satisfaction of a stop condition. Despite the simplicity of this technique, the first detected error can reinforce thereby deteriorating the quality of the final result. Moreover, the convergence is not always guaranteed for complex data. This technique has been successfully applied to face recognition systems using Principal Component Analysis (PCA) and fingerprint recognition using the method of measurement strings [5] [6] [7] [8] [9].

• The semi-supervised learning is also called S3VM SVMT (transductive SVM). It extends the standard SVM with unlabeled data. Its goal is to find the data labels so that the hyperplane has a maximum margin on the original labeled and unlabeled data [4] [10] data. So SVMT is based on a clear and applicable mathematical foundation where SVM is applicable [11]. However, there are some optimization difficulties to find the exact solution of SVMT.

• The generative models [12] [13] are based on statistical models. The most commonly used models are: the Gaussian mixture (GMM) for image classification, the mixture of multinomial distributions (Naïve Bayes) for categorization of text and hidden Markov model (HMM) for speech recognition. Note that it is often difficult to verify the accuracy of the model used. Therefore, unlabeled data can be misclassified if the selected generative model is wrong.

• Co-learning (Co-training) [14] is a version of the self-learning adapted to the use of at least two classifiers that will help each other to improve each other. It uses a second classifier that allows the inclusion of data with high variability that was not accepted by the first classifier [15] [16] [17] [18] [19] [20]. It is based on the assumption of the existence of two independent projections of the same data space. Two classifiers are trained on these two projections must labeled identically the same data. This technique, used in many studies [21] [22] [23] [4], has the advantage of being less sensitive compared to the self-learning errors.

The examination of various semi-supervised and previous work based on SSL learning techniques allowed us to choose the Co-training. Indeed, it not only has the advantage that it is less sensitive to noise, robust and fast, but also its theoretical basis is promising for complex contexts.

## 3. Co-training

In this section, we start by presenting basic steps of the classical variant of co-training, and then we present the proposed new variant.

### 3.1 Classical variant of Co-Training

Classical Co-Training steps are resumed in Algorithm 1 and illustrated by figure 1.

**Algorithm 1 : Classical Co-training**

```
Inputs
L: a small set of manually labeled data (M)
U: a large set of unlabeled data
V1 and V2: features vectors of U
Outputs
C1 and C2: two classifiers
L: initial L expanded by automatically labeled data
1. For i=1 to N
    {
     1.1. Use L for learning by h1 using V1
     1.2. Use L for learning by h2 using V2
     1.3. Classify and label data of U by h1 and h2
     1.4. Select the best labeled data from U
     1.5. and add them to L
   }
2. Generate C1 and C2 from L
3. Delete L
4. Classify data of U by C1 and C2
5.Label data according to the most confidante classification
```

The Co-training algorithm requires two learning techniques ($h1$ and $h2$) and two subsets of different features related to the labeled data ($L$). The principle of the algorithm [14] [24] requires that (i) the features can be partitioned into two sets, (ii) each subset of features set is used for learning $h1$ or $h2$, and (iii ) the two subset are conditionally independent [4]. In each iteration, unlabeled samples ($U$ ) are labeled by each learning technique (line 1.3., Algorithm 1). Based on a confidence score, the best classified samples are selected and added to the labeled base (lines 1.4. and 1.5., Algorithm 1). Once the two classifiers ($C1$ and $C2$) are obtained (*i.e*, after $N$ iterations) (line 2, Algorithm 1), the training base $L$ is deleted (lines 3 Algorithm 1). Labels of new instances are predicted by the most confident classifier on the sample (lines 4 and 5, Algorithm 1).

Despite the success of the classic variant of co-training in the classification of different types of data (see [25] [26] [27] [28] [29] [30]), this variant does not solve robustly the problem of classification of the complex data, in particular those who present strong similarity. In fact, when classical variant of Co-training is applied on data, two classifiers can be generated from the learning set L extended after N iterations. However, the learning set is not representative of the reality of complex data whose features change frequently. In addition, this set is regenerated for each new entry (unlabeled data). This constraint has led us to propose a new variant of co-training suitable for better complex data classification.

### 3.2 New variant of du Co-training

The objective of the new variant of Co-training is to generate a classifier adapted to the complexity of the data.

Unlike the classic variant of co-training, the proposed variant has three advantages: (1) it generates a classifier more generic by a learning technique from a significant base automatically labeled and not from the base $L$ extended after $N$ iterations, this database is called MGLP (Model Generation Labeled Pool), (2) selects a small labeled set of $L$ from MGLP data, and (3) increasing the level of selection confidence of the best classified samples through a prediction model obtained by a supervised learning. Note that the proposed variant of Co-training led to an adaptation to complex data with an optimal setting. The process

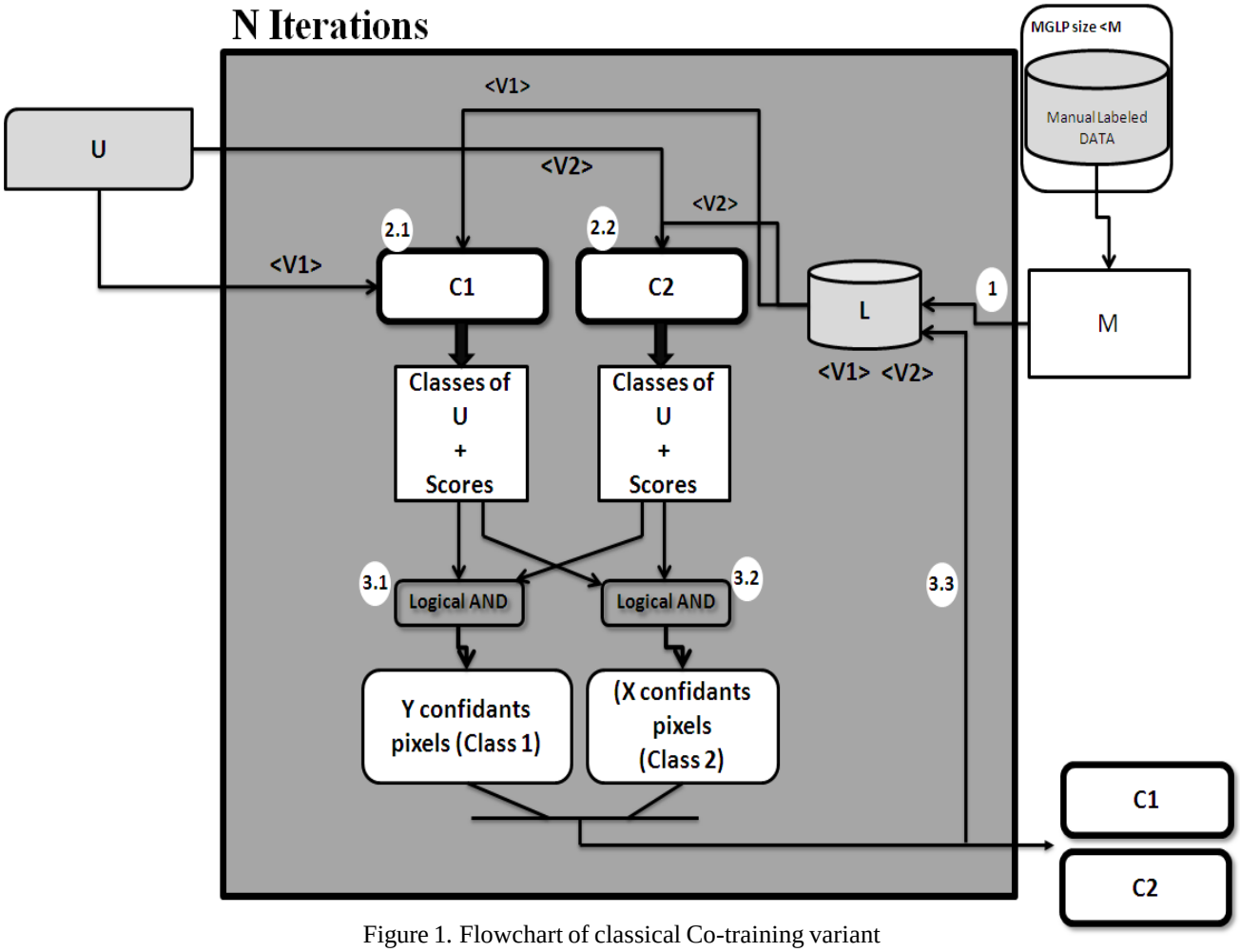



Figure 1. Flowchart of classical Co-training variant

of the new variant of co-training is shown in Figure 2. The first stage builds the basis MGLP and the second step generates the classifier.

In the construction phase of MGLP (stage 1, Figure 2), a learning step by two classifiers ($C1$ and $C2$) is iterated $N$ times for each new data. Note that the fixation of an optimal number $N$ of iterations is necessary for better speed of MGLP construction. Learning by $C1$ and $C2$ is performed respectively on two feature vectors ($V1$ and $V2$) of small labeled data called $L$. Reducing the size of the training set is required to minimize the learning time. Thus, it is important to ensure that this learning base contains a set of labeled samples quite significant. For this, the initial $M$ samples of $L$ are randomly selected from a database of manually labeled samples. Once the size of the MGLP is greater than $M$, the initial samples of $L$ will be selected from MGLP.

In every learning iteration, the most confidants labeled data of each class are added to the database $L$ and MGLP. The selection of the most confidants data is performed more reliable by taking into account the decision of classification given by a prediction model obtained offline by a supervised learning.

After constructing the learning base labeled automatically MGLP, the objective of the second step is to generate online, a classifier for the classification of complex data. Therefore, a learning technique ($T$) is used which have provided, for this task, efficient and fast results.

**4. Experimental Evaluation**

The choice of an area of application for this new variant of co-training is crucial for evaluation. We note that this variant as well

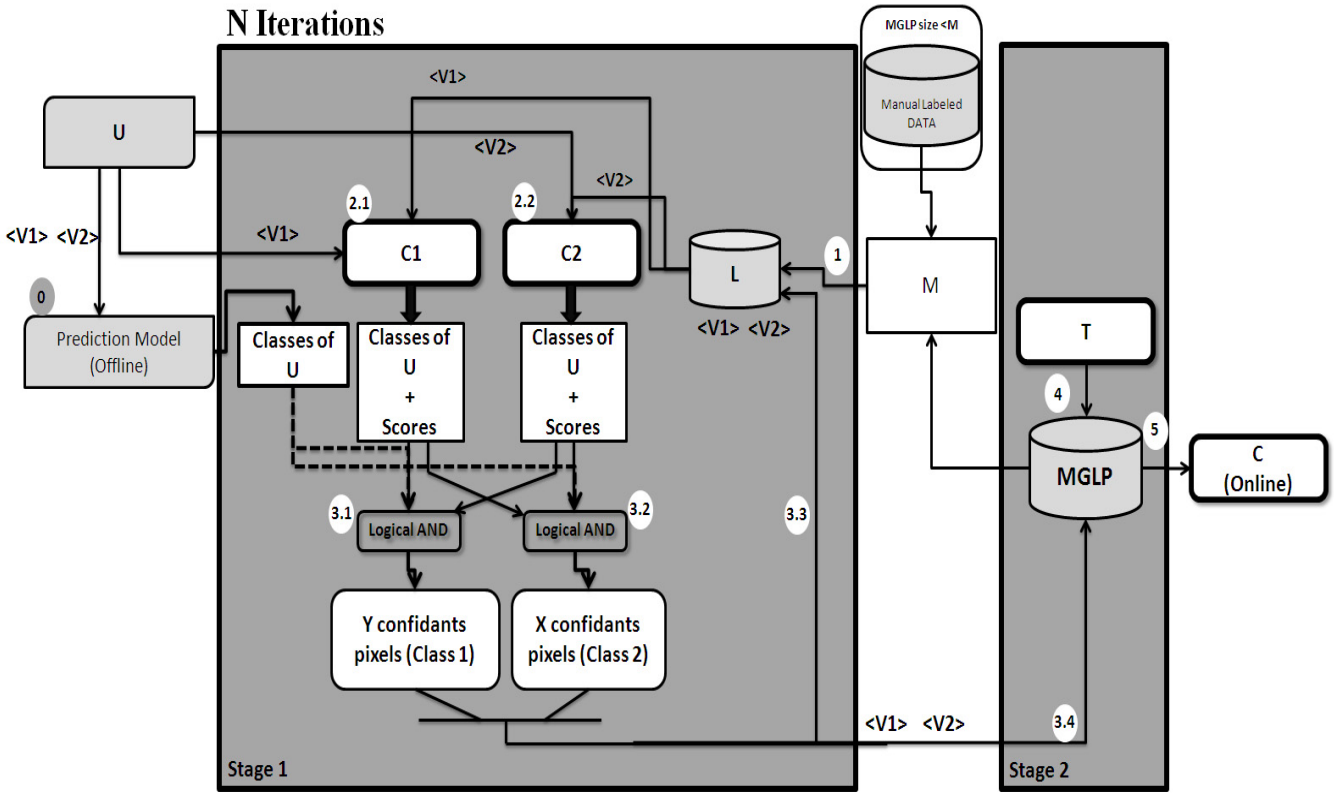



Figure 2. Flowchart of the new variant of Co-training

as the classic variant were adopted respectively by [32] and [33] for the online classification of moving pixels in shadow / non-shadow from a video stream. Indeed, the detection and removal of shadows related to moving objects show great interest in computer vision applications. The removal of moving shadows increases the accuracy of detecting moving objects, thus increasing the effectiveness of these applications. The strong similarity between the characteristics of moving objects and shadows pixels makes their classification more complex. In addition, their characteristics change frequently with the dynamic conditions of the observed scenes.

In the same context (online classification of moving pixels in shadow / non-shadow from a video stream), learning in both variants, is performed by two SVM ($C1 = SVM\ 1$ and $C2 = SVM\ 2$ ) with a radial basis kernel (RBK). Setting the classical variant has been validated by an experimental study [33]. This study has established for the classical variant, $N = 20$ iterations and $M = 100$ pixels (the original pixels of $L$).

In the new variant of co-training the classifier $C$ is generated by a $T = C4.5$. Remember that this new variant uses a prediction model with $C1$ and $C2$. This model, called *PM_Shadow*, is the prerequisite result of considerable offline work. *PM_Shadow* is obtained by a supervised learning respecting a series of steps in order to generate a more generic model. We find it interesting to present these steps in subsection *A*. In addition, we have conducted a series of experiments, described in subsection B, which aim to determine the optimal number of iterations $N$ and the most appropriate initial size $M$ of the database $L$.

In subsection C, we present and discuss the quantitative results obtained by the two variants for the online classification of moving pixels in shadow/non-shadow from a video stream.

**4.1 PM_Shadow generation**

In the first step, we built a learning base from the most famous sequences in the literature. These sequences are acquired under typical conditions (outside or inside) and include different objects classes and different noise levels. Manual labeling performed resulted in a corpus composed of 1.897.998 shadow pixels, 1.791.562 non-shadow pixels. In this step, the choice of the most

ID3

Learning Set

| | A | B |
|---|---|---|
| A | 93,62% | 6,38% |
| B | 2,67% | 97,32% |
| CDR | 95,76% | |
| FDR | 4,24% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 90,25% | 9,75% |
| B | 4,95% | 95,05% |
| CDR | 93,02% | |
| FDR | 6,98% | |

Figure 3. CDR and FDR obtained by ID3 for learning Set and Inactive Example

C4.5 (1993)

Learning Set

| | A | B |
|---|---|---|
| A | 93,2% | 6,8% |
| B | 2,36% | 97,64% |
| CDR | 95,77% | |
| FDR | 4,23% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 90,25% | 9,75% |
| B | 4,3% | 95,7% |
| CDR | 93,4% | |
| FDR | 6,6% | |

Figure 4. CDR and FDR obtained by C4.5 (1993) for learning Set and Inactive Example

Cost-Sensitive C4.5(2001)

Learning Set

| | A | B |
|---|---|---|
| A | 93,03% | 6,97% |
| B | 2,47% | 97,53% |
| CDR | 95,64% | |
| FDR | 4,36% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 90,23% | 9,77% |
| B | 4,3% | 95,7% |
| CDR | 93,39% | |
| FDR | 6,61% | |

Figure 5. CDR and FDR obtained by Cost-Sensitive C4.5 (2001) for learning Set and Inactive Example

relevant shadow features is crucial. It determines the quality of established models. Therefore, we identified the more relevant shadows features (see [31]). These features are calculated for each pixel.

Following the preparation of the training data, we carried out various experiments to find the most accurate prediction model. This stage consists to extract useful knowledge from the training data set. However, in the literature, there are several techniques

One-Vs-All-Decision-arbre

Learning Set

| | A | B |
|---|---|---|
| A | 90,53% | 9,47% |
| B | 36,52% | 63,48% |
| CDR | 74,87% | |
| FDR | 25,13% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 90,49% | 9,51% |
| B | 36,6% | 63,4% |
| CDR | 74,84% | |
| FDR | 25,16% | |

Figure 6. CDR and FDR obtained by One-Vs-All-Decision-arbre for learning Set and Inactive Example

A limited search:

Learning Set

| | A | B |
|---|---|---|
| A | 76,94% | 23,06% |
| B | 5,34% | 94,66% |
| CDR | 87,2% | |
| FDR | 12,8% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 76,86% | 23,14% |
| B | 5,28% | 94,72% |
| CDR | 87,18% | |
| FDR | 12,82% | |

Figure 7. CDR and FDR obtained by A limited search for learning Set and Inactive Example

Improved Chaid:

Learning Set

| | A | B |
|---|---|---|
| A | 80,33% | 19,67% |
| B | 9,4% | 90,6% |
| CDR | 86,27% | |
| FDR | 13,73% | |

Inactive Example

| | A | B |
|---|---|---|
| A | 80,27% | 19,73% |
| B | 9,32% | 90,68% |
| CDR | 86,28% | |
| FDR | 13,72% | |

Figure 8. CDR and FDR obtained by Improved Chaid for learning Set and Inactive Example

of supervised learning. Among the most important criteria of theoretical comparison of these techniques is the understandability of produced prediction model. Based on this criterion, we decided to use decision trees. These latter focuses on usability and readability results. They also produce results easily interpretable and therefore exploitable, and presented in the form of logical classification rules.

We have studied six techniques based on decision trees including: *ID*3, *C*4.5, *Cost-Sensitive C*4.5, *One-Vs-All Decision Tree*, *A limited search* and *Improved Chaid*. We divided our database into training data set (Learning Set) (70%) and test data set (Inactive Example) (30%). We computed, for each technique, the rate of correct classification (Correct Detection Rate (CDR)) and the rate of incorrect classification (False Detection Rate (FDR)), the additional (TCD) based on the confusion matrix. As shown in Figures 3, 4, 5, 6, 7, and 8, the best rates of CDR and FDR are given by C4.5 1993.

**4.2 Proposed setting for the new Co-training variant (M and N)**

For different values of *N* and *M*, we calculated the shadow detection accuracy (*h*) (equation 1), shadow discrimination accuracy (x) (Equation 2), the correct classification rate (Classification Accuracy (*CA*)) (Equation 3) and the execution time (s).

$$\eta = \frac{TP_s}{TP_s + FN_s}$$

$$\xi = \frac{TP_F}{TP_F + FN_F}$$

$$CA = \frac{TP_s + TP_s}{TP_F + FN_F + TP_s + FN_s}$$

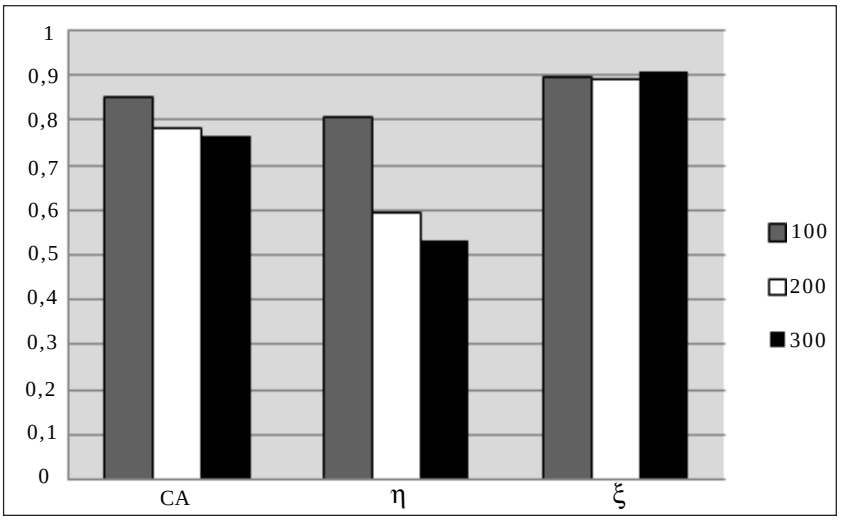



Figure 9. Average rates of *CA*, $\xi$ and $\eta$ obtained for differents initial sizes ($M$ ={100, 200, 300}) of $L$ ($N$ = 5)

With:

- *S* and *F* denote the pixel shadow and non-shadow pixels.
- $TP_S$: the number of pixels belonging to the class of shadow pixels which is affected by the class of shadow pixels.
- $FN_F$: the number of pixels belonging to the class of non-shadow pixels which is affected by the class of shadow pixels.
- $FN_S$: the number of pixels belonging to the class of shadow pixels which is affected by the class of non-shadow pixels.
- $TP_F$: the number of pixels belonging to the class of non-shadow pixels which is affected by the class of non-shadow pixels.

According to our intensive experiments, we have chose the number of iteration N = 5 and the initial size M = 100 pixels. Some results of these experiments are presented in Figures 9, 10, 11 and 12.

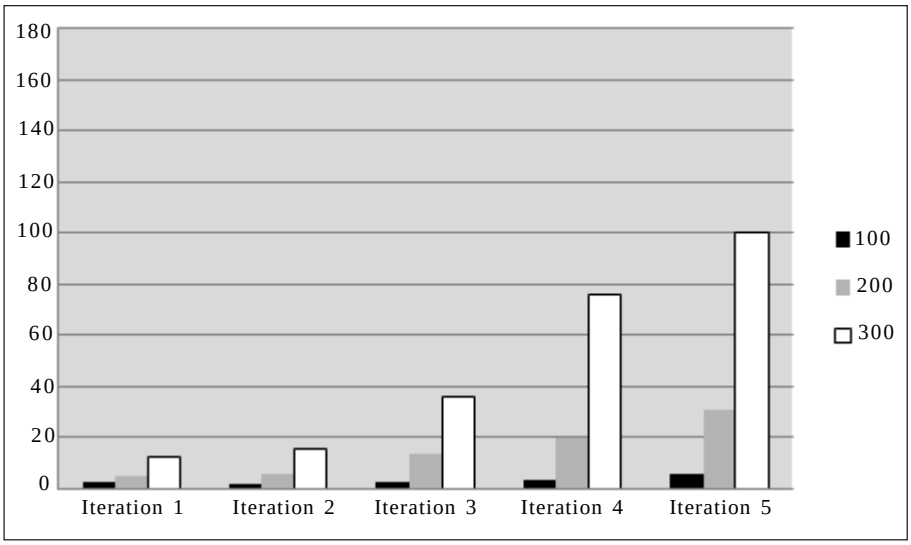



Figure 10. Execution times of five iterations with differents initial sizes ($M$ ={100, 200, 300}) of $L$ ($N$ = 5)

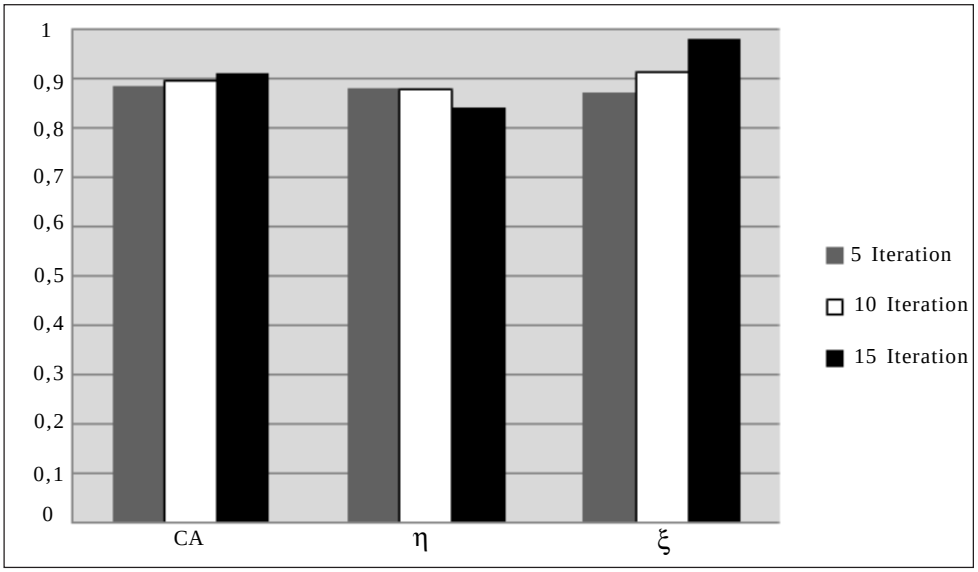



Figure 11. Average rates of *CA*, $\eta$ and $\xi$ obtained for differents number of iterations ($N$ ={5,10,15}) with $M$ =100

Figures 9 and 10 show, respectively, for different sizes of $M$ (100, 200 and 300 pixels), the average rates of *CA*, η and *x*, and the execution time required to perform learning and generation of $C1$ and $C2$ by two SVM for five iterations. Recall that in each iteration, the initial size ($M$) of $L$ increases by the more confidant classified pixels. The best rate of *CA*, η and ξ (≈ 83%) are given by $M$ = 100. In addition, for the same value we obtained an optimal execution time for five iterations.

Figures 11 and 12 show, respectively, for different numbers of iterations $N$ (5, 15 and 20), the average rate of *CA*, η and ξ, and execution time required for learning and generation of $C1$ and $C2$ by two SVM for $M$ = 100. For $N$ = 5, the proposed variant records *CA*, η and ξ rates (≈ 88%) close to those reported for $N$ = 10 (≈ 89%) and $N$ = 15 (≈ 91%) with a very significant gain in terms of execution time.

Obviously, the selection of M pixels from MGLP and the presence of a prediction model that increases the level of confidence

assure both the optimality and robustness of our variant.

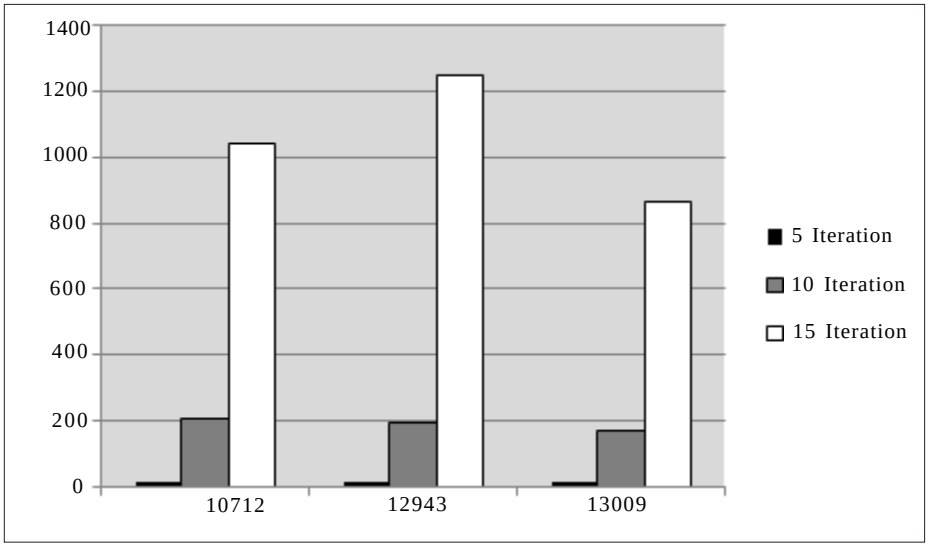



Figure 12. Execution times of differents number of iterations ($N$ ={5, 10 and 15}) ($M$ =100)

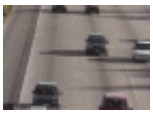
| | ***Highway I***  Outside | ***IntelligentRoom***  Inside |
|---|---|---|
| **Light Source** | Natural | MultipleArtificial |
| **Speed of Moving Object** | High speed (Vehicles) | Low speed (Person) |
| **Shadow Shape** | Big shape | Small shape |
| **Shadow Color** | Dark | Clear |
| **Texture level in Background were Shadow is projected** | Low texture | Varied Texture |

Table 1. Corpus d'évaluation quantitative

**4.3 Comparative study**

In order to validate the robustness and performance of our co-training variant, the comparative study was performed within two series of experiments on sequences[1] *HighwayI* and *IntelligentRoom* (see Table I). The first experiment is a comparison between the results of classification of moving pixels in sha dow/non-shadow given by the new variant of co-training and the classical one. Quantitative scores (shadow detection accuracy ($\eta$) and shadow discrimination accuracy ($\xi$)) were used for this evaluation. The second experiment is a comparison in terms of execution time ($s$) between the two variants.

The comparative study (see Figure 13) with a classical variant [33] shows that the results achieved by the new variant exceed those recorded by the classical variant. Indeed, our variant recorded the best recall rate on the detection of moving pixels and the detection of shadow pixels ($\eta$ and $\xi$ between 84.85% and 93.47%). In addition, the classical variant [33] gives rates $\eta$ and $\xi$ between 81.23% and 89.76%. The gains in our variant are due, on the one hand, to the adaptive nature of our variant, and on the other hand, to the independence of our variant of the diversity and dynamism of the environment. This further proves the improvement reported by learning and generation of a classifier from a significant base MGLP.

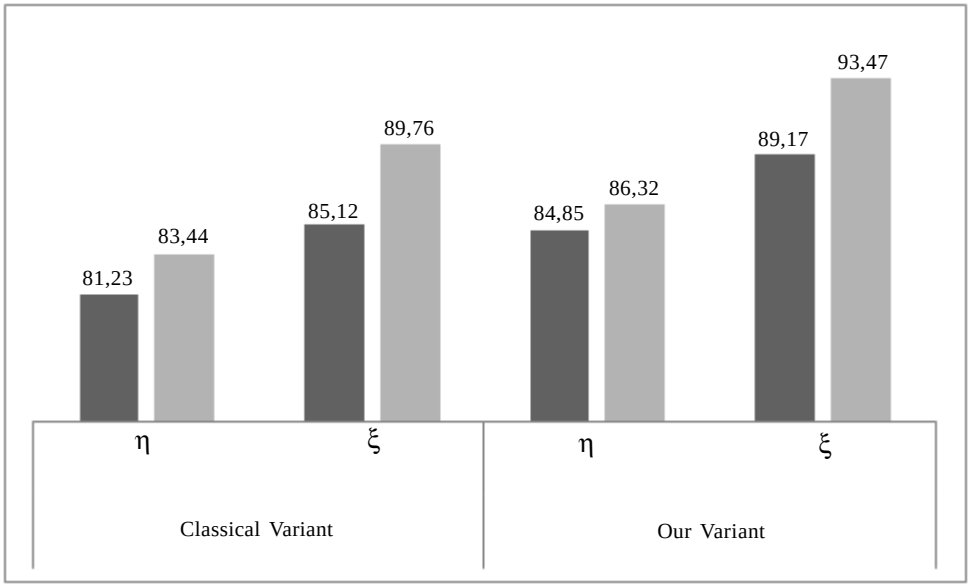
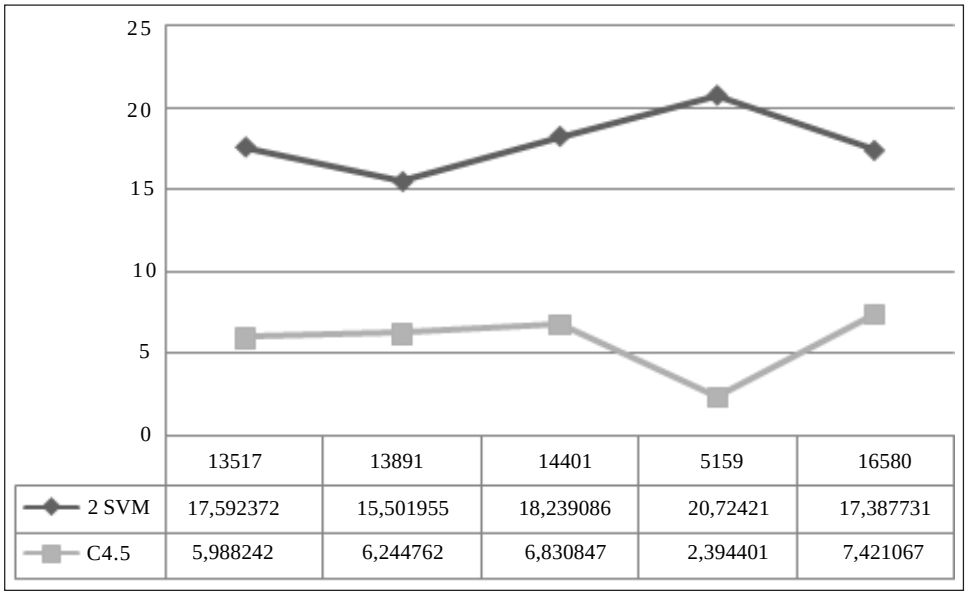



Figure 13. Average rates of $\xi$ and $\eta$ obtained by our variant (light grey) and classical variant (darken grey) on *HighwayI* (right) and *IntelligentRoom* (left)




| | 13517 | 13891 | 14401 | 5159 | 16580 |
|---|---|---|---|---|---|
| 2 SVM | 17,592372 | 15,501955 | 18,239086 | 20,72421 | 17,387731 |
| C4.5 | 5,988242 | 6,244762 | 6,830847 | 2,394401 | 7,421067 |

Figure 14. Execution time for learning and generating final classifiers by our variant (with C4.5) and classical variant (with 2SVM) of Co-training

Figure 14 shows the comparison results between our variant of Co-training ($N$ = 5 and $M$ = 100) and the classic variant ($N$ = 20 and $M$ = 100) in terms of execution time for learning and generating classifiers (respectively $C$ and ($C1$ and $C2$)) for several sets of data (13517, 13891, 14401, 16580 and 5159 pixels). This comparison shows that the classical variant requires much more time than our variant for learning and generating a classifier from a large data set (after 20 iterations). Indeed, the learning principle by SVM prevents rapid convergence, *i.e* the separating hyperplane is hardly found in a large complex database.

**5. Conclusion**

In this paper, we propose a Semi-Supervised Learning by Co-training approach. In fact, this proposal comes from the study of classical variant which is not suitable for complex data.

The proposed variant is based on a varied database (MGLP) online labeled automatically, and generates a prediction model from this base.

The construction of MGLP relies on results of a considerable offline work to generate a prediction model. This offline work will not have to be redone; their results (prediction model) can be used to build another database MGLP.

Performances of this new variant have been demonstrated through a quantitative evaluation. The quantitative results presented in this paper show improvements in terms of rate and time. The obtained results show the independence of our variant of the diversity and dynamism of the environment. The various experiments performed are leaving a promising exploitable in the online classification of different types of complex data.

**References**


[1] Kristin P. Bennett, Ayhan Demiriz. (1999). Semi-supervised support vector machines. *In*: Proceedings of the conference on Advances in neural information processing systems, p. 368–374.

[2] Greg Schohn, David Cohn. (2000). Less is more: Active learning with support vector machines. *In*: Proceedings of the Seventeenth In ternational Conference on Machine Learning, p. 839–846.

[3] Andre Guggenberger. (2008). Semi-supervised learning with support vector machines. *Technical report*.

[4] Xiaojin Zhu. (2008). Semi-supervised learning literature survey. *Technical report*, New York.

[5] Giot, Romain. (2012). Contributions à la dynamique de frappe au clavier : multibiométrie, biométrie douce et mise à jour de la référence, Université de Caen.

[6] Marcialis, G., Rattani, A., Roli. F. (2008). Biometric template update: an experimental investigation on the relationship between update errors and performance degradation in face verification. *In*: SSPR&SPR, p. 684–693. Springer.

[7] Ajita Rattani. (2010). Adaptive Biometric System based on Template Update Procedures. Thèse de doctorat, Dept. of Electrical and Electronic Engineering University of Cagliari.

[8] Roli, F., Marcialis, G. (2006). Semi-supervised pca-based face recognition using self-training. *In*: Structural, Syntactic, and Statistical Pattern Recognition, p. 560–568. Springer.

[9] Fabio Roli, Luca Didaci, Gian Luca Marcialis. (2008). Advances in Biometrics, chapitre Adaptive Biometric Systems That Can Improve with Use, p. 447–471. SpringerLink.

[10] Chapelle, O., Sindhwani, V., Keerthi, S. (2008). Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research*, 9, 203-233.

[11] Tian, Xilan. (2012). Apprentissage and Noyau pour les Interfaces Cerveau-machine. Diss. INSA de Rouen.

[12] Nigam, K., McCallum, A. K., Thrun, S., Mitchell, T. (2000). Text classification from labeled and unlabeled documents using em. Machine Learning, p. 103–134.

[13] Fujino, A., Ueda, N., Saito, K. A hybrid generative/discriminative approach to semi-supervised classifier design. *The Twentieth National Conference on Artificial Intelligence*.

[14] Avrim Blum, Tom Mitchell. (1998). Combining labeled and unlabeled data with co-training. *In*: Proceedings of the eleventh annual conference on Computational learning theory, p. 92–100.

[15] Didaci, L., Marcialis, G., Roli, F. (2009). Modelling frr of biometric verification systems using the template coupdate algorithm. *In*: Lecture Notes in Computer Sciences, ICB.

[16] Ajita Rattani. (2010). Adaptive Biometric System based on Template Update Procedures. Thèse de doctorat, Dept. of Electrical and Electronic Engineering University of Cagliari.

[17] Ajita Rattani, Gian Luca Marcialis, Fabio Roli. (2008). Boosting gallery representativeness by co-updating face and fingerprint verification systems. 5th Summer School for Advanced Studies on Biometrics for Secure Authentication.

[18] Ajita Rattani, Gian Luca Marcialis, Fabio Roli. (2008). Capturing large intra-class variations of biometric data by template co-updating. *In*: Computer Vision and Pattern Recognition Workshops. CVPRW'08. *IEEE Computer Society Conference on* 2008c.

[19] Roli, F., Didaci, L., Marcialis, G. (2007). Template co-update in multimodal biometric systems. *In*: International Conference on Biometrics (ICB), p. 1194–1202. Springer.

[20] Fabio Roli, Luca Didaci, Gian Luca Marcialis. (2008). Advances in Biometrics, chapitre Adaptive Biometric Systems That Can Improve with Use, p. 447–471. SpringerLink.

[21] Jones, R. (2005). Learning to extract entities from labeled and unlabeled text (technical report cmu-lti-05-191). *Technical report*.

[22] Zhou, Z. -H., Zhan, D. -C., Yang, Q. (2007). Semi-supervised learning with very few labeled training examples. Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07).

[23] Balcan, M. -F., Blum, A. (2006). An augmented pac model for semi-supervised learning. *In*: Semi-Supervised Learning, p. 397–420, 226.

[24] Mitchell, T. (1999). The role of unlabeled data in supervised learning. *In*: Proceedings of the Sixth International Colloquium on Cognitive Science.

[25] Javed, O., Ali, S., Shah, M. (2005). Online detection and classification of moving objects using progres-sively improving detectors. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 696–701.

[26] Nair, V., Clark, J. J. (2004). An unsupervised, online learning framework for moving object detection. *In*: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 317–324.

[27] Ferencz, A., Learned-Miller, E.G., Malik, J. (2005). Building a classification cascade for visual identification from one example. Tenth IEEE International Conference on Computer Vision, 1, 286–293.

[28] Hewitt, R., Belongie, S. (2006). Active learning in face recognition: Using tracking to build a face model. Conference on Computer Vision and Pattern Recognition Workshop, CVPRW '06., p. 157–157.

[29] Feng Tang, Brennan, S., Qi Zhao, Hai Tao. (2007). Co-tracking using semi-supervised support vector machines. IEEE 11th International Conference on Computer Vision, ICCV, p. 1–8.

[30] Jepson, A. D., Fleet, D. J., El-Maraghi, T. F. (2003). Robust online appearance models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1296–1311,.

[31] Salma, Kammoun Jarraya. (2013). Détection par Modélisation en-ligne du fond and suivi d'objets en mouvement dans un flux vidéo capté par une caméra fixe. Université de Sfax.

[32] Salma Kammoun Jarraya, Rania Rebai Boukhris, Mohamed Hammami, Hanêne Ben-Abdallah. (2012). Cast shadow detection based on semi-supervised learning. *In*: International Conference on Image Analysis and Recognition, p. 406–414. Springer LNCS.

[33] Joshi, A. J., Papanikolopoulos, N. P. (2008). Learning to detect moving shadows in dynamic environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (11) 2055–2063.