Sparse and Non-Sparse Multiple Kernel Learning for Recognition

Mitchel Alioscha-Pérez[1], Hichem Sahli[2], Isabel González[2], and Alberto Taboada-Crispi[1]

[1] Centro de Estudios de Electrónica y Tecnologias de la Información (CEETI),
Universidad Central de Las Villas, Villa Clara,
Cuba

[2] Department of Electronics and Informatics (ETRO),
Vrije Universiteit Brussel, Brussels,
Belgium

sirmichel@gmail.com, {hichem.sahli, igonzale@etro.vub.ac.be}, ataboada@uclv.edu.cu

Abstract. The development of Multiple Kernel Techniques has become of particular interest for machine learning researchers in Computer Vision topics like image processing, object classification, and object state recognition. Sparsity-inducing norms along with non-sparse formulations promote different degrees of sparsity at the kernel coefficient level, at the same time permitting non-sparse combination within each individual kernel. This makes MKL models very suitable for different problems, allowing adequate selection of the regularizer according to different norms and the nature of the problem. We formulate and discuss MKL regularizations and optimization approaches, as well as demonstrate MKL effectiveness compared to the state-of-the-art SVM models using a Computer Vision Recognition problem.

Keywords. Multiple kernel learning, object state recognition, norm regularizers, analytical updates, cutting plane method, Newton's method.

Aprendizaje de múltiples núcleos esparcidos y no esparcidos para reconocimiento

Resumen. El desarrollo de técnicas MKL (aprendizaje de múltiples núcleos) ha sido de particular interés para los investigadores en el aprendizaje automatizado, en tópicos de visión por computadora, así como para el procesamiento de imágenes, clasificación de objetos, y reconocimiento del estado de los objetos. En ecuaciones donde se combinan múltiples núcleos, las normas de inducción de dispersión junto con las formulaciones de no dispersión, promueven diferentes grados de dispersión a nivel de los coeficientes de combinación, mientras que permiten la combinación no esparcida en los núcleos individuales. Esto hace de los modelos MKL muy adecuados para diferentes problemas, permitiendo la selección óptima del regularizador, y así lograr un mejor reconocimiento de acuerdo a la naturaleza del problema. En este trabajo, formulamos y discutimos las diferentes regularizaciones de MKL y los métodos de optimización relacionados, demostrando su efectividad en un problema de reconocimiento de visión por computadora.

Palabras clave. Aprendizaje de múltiples núcleos, reconocimiento del estado de objetos, regularizadores de normas, actualizaciones analíticas, método de planos cortantes, método de Newton.

1 Introduction

The use of kernel methods has taken a considerable importance in Computer Vision problems. Among large margin methods, SVM (support vector machine) is one of the most commonly used techniques. Indeed, the proven generalization ability of SVM (and specifically, of binary classifiers) in regression and classification problems involved this model very often in object classification and recognition. The drawback of using SVM in classification problems is that the kernel should be specified (given) and this is often done in an empirical way. This problem is circumvented by using the MKL framework, allowing a joint learning of the optimal kernel mixture and the alpha coefficients from a list of candidate kernels. This automatic selection usually results in much more interpretable and/or accurate models. The mixture selection is very close to the choice of the regularizer of the MKL formulation. The chosen regularizer will lead to the degree of sparsity that will be present in the

resulting kernel mixture (at the kernel mixture coefficient level). Therefore, depending on the nature of the problem, by choosing different regularizers, we will get different kernel mixtures, each one with potentially different accuracy results for the overall classification process. In this work, we discuss the use of MKL in Computer Vision recognition problems without considering any particular group structure or prior information on the problem.

Recent formulations of MKL models allow efficient computation of resulting kernel mixtures in different arbitrary ℓ_p norms, considering only $p \geq 1$. In this work, we will separate discussions of the MKL formulation into sparse $p = 1$ and $1 < p < \infty$ non-sparse formulations in order to investigate the importance of sparsity on the kernel coefficient in different Computer Vision recognition problems. Here, our discussion will be limited to the binary classification problem. In this work, we compare different sparsity degree selection, along with different optimization strategies and algorithms. Specifically, for the purpose of this study, we implemented analytical update methods such as Reduced Gradient and Newton's method with Linear Search methods such as Golden and Fibonacci search, and Cutting Plane methods based on Semi-Infinite Programming (SIP). Finally, we perform an analysis of the obtained results and present our conclusions.

2 MKL Framework

Multiple Kernel Learning formulation was first proposed by [9] as a convex combination of kernels projected on the cone defined by a set of semi-definite kernel matrices. Later, Bach *et al.* [1] established its equivalence with a formulation where the kernel mixture coefficients were regularized in a mixed $\ell_1 - \ell_2$ norm; this formulation is the basis of actual MKL models. The MKL comes from the large margin methods, more precisely, from SVM, and consists in finding, given a set of observations $(x_1, y_1)..(x_n, y_n) \in \mathcal{X} \times \mathcal{Y}$ with $x = (x_1, .., x_n)$, a hypothesis $f :\rightarrow \mathcal{X}, f \in \mathcal{H}$ that best generalizes the unseen data. It is sufficient to define a symmetric semi-positive definite kernel $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{R}$ as our reproducing kernel, to inherit its corresponding and unique reproducing kernel Hilbert space $\mathcal{H}$ (rkHs) (see Moore-Aronzajm theorem) with $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{R}$. For SVM, the search in the primal formulation of f is defined as

$$\begin{aligned} \min_{f \in \mathcal{H}, b, \xi_i \geq 0} \quad & \tfrac{1}{2} \|f\|_{\mathcal{H}}^2 + C \sum_{i=1}^{n} \xi_i \\ s.t. \quad & \forall_i y_i \left(f(x_i) + b \right) \geq 1 - \xi_i \end{aligned} \tag{1}$$

On the dual derivation of the previous, we use $f(x) = \langle f(.), k(., x) \rangle_{\mathcal{H}}$, which is the reproducing kernel property, to substitute $\langle k(., x_i), k(., x_j) \rangle_{\mathcal{H}} = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} = k(x_i, x_j)$ where $\phi : \mathcal{X} \rightarrow \mathcal{H}$ map features from input space x to Hilbertian space $\mathcal{H}$, obtaining the following dual formulation of Eq. 1:

$$\begin{aligned} \max_{\alpha_i \in \mathbb{R}} & \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j k(x_i, x_j) \\ \text{s.t.} \quad & 0 \leq \alpha \leq C \quad , \quad \sum \alpha_i y_i = 0 \ \forall_i \end{aligned} \tag{2}$$

Specifically in the MKL context, not only a single hypothesis $f \in \mathcal{H}$, but several hypotheses $f_m \in \mathcal{H}_m$ will be found, where each kernel function $k_m(x, x')$ is used to compute inner products in $\mathcal{H}_m$, with $\phi_m : \mathcal{X} \rightarrow \mathcal{H}_m$ as a feature map, considering $f_m = d_m f_m' \in \mathcal{H}_m$ and $\mathcal{H} = \bigoplus_{i=1}^{n} \mathcal{H}_m$. Then we can state that

$$k(x, x') = \sum_{m=1}^{M} d_m k_m(x, x')$$

and $\mathcal{H}$ is its respective rkHs; therefore, the formulation in [1] for the mixed norm $\ell_1 - \ell_2$ can be defined as

$$\begin{aligned} &\min_{\{f_m\},b,\xi_i,d} \ \tfrac{1}{2}\sum_{m=1}^{M}\left(d_m\|f_m\|_{\mathcal{H}_m}\right)^2+C\sum_{i=1}^{n}\xi_i \\ &s.t.\forall_i\, y_i\left(\sum_{m=1}^{M}\langle f_m,\phi_m(x)\rangle_{\mathcal{H}_m}+b\right)\geq 1-\xi_i \\ &w.r.t. \quad d_m\in\mathbb{R}+,\ \sum_{m=1}^{M}d_m=1\ \ \xi_i\in\mathbb{R}_+ \\ &\qquad b\in\mathbb{R} \quad f_m\in\mathcal{H}_m \end{aligned} \tag{3}$$

The ℓ_1 norm will lead to sparse solutions on $\boldsymbol{d}=(d_1,..,d_m)$ while the ℓ_2 norm keeps non-sparse combinations of the kernel function per selected model $f_m\in\mathcal{H}_m$. It is important to remark that the previous formulation is convex over the simplex resulting of the ℓ_1 norm constraint. However, it is non-smooth; for instance, it is not differentiable for $d_m=0$. Although [9] has a nice property of inducing sparsity on the resulting kernel's coefficients, it is not differentiable, and hence, limiting the set of optimization strategies to be used for solving it. In [1], the author proposed a smoothing technique by introducing Moreau-Yosida terms on the objective function in order to make it smooth and then applied Second Order Cone Programming (SOCP) for its dual derivation. As a result, it was possible to use Sequential Minimal Optimization (SMO) techniques which allowed dealing with medium scale sized problems.

2.1 Sparsity Inducing Norms

The non-smoothness of Eq. 3 has been overcome in a latter formulation [10] where the differentiability conditions for the objective function are granted:

$$\begin{aligned} &\min_{\{f_m\},b,\xi_i,d} \ \tfrac{1}{2}\sum_{m=1}^{M}\tfrac{1}{d_m}\|f_m\|^2_{\mathcal{H}_m}+C\sum_{i=1}^{n}\xi_i \\ &s.t.\forall_i\, y_i\left(\sum_{m=1}^{M}\langle f_m,\phi_m(x)\rangle_{\mathcal{H}_m}+b\right)\geq 1-\xi_i \\ &w.r.t. \quad d_m\in\mathbb{R}+,\ \sum_{m=1}^{M}d_m=1\ \ \xi_i\in\mathbb{R}_+ \\ &\qquad b\in\mathbb{R} \quad f_m\in\mathcal{H}_m \end{aligned} \tag{4}$$

Above, a framework is convened where $\frac{x}{0}=0$ if $x=0$ and ∞ otherwise, reducing the space $\mathcal{H}_m$ to the space of those hypothesis f_m such as $\frac{1}{d_m}\|f_m\|_{\mathcal{H}_m}<\infty$, which will result in the fact that $f_{m=0}$ whenever $d_m=0$ in order to reach the finite objective value; the mixed $\ell_1-\ell_2$ norm of Eq. 3 is replaced by a weighted ℓ_2 norm in the previous Eq. 4, and its equivalence with Eq. 3 is established by making use of the Cauchy-Schwartz inequality (see the proof in [10]). This smooth formulation allowed the use of analytical update methods, mostly gradient-based ones, in order to compute the descent direction during the optimization process. Some other dual derivations lead to Semi-Infinite Programming (SIP), more precisely, to Semi-Infinite Linear Programming (SILP) for Eq. 3, used to continuously find upper bound solutions by generating many linear constraints. This is known as the cutting plane method [12] (note that this method does not require the differentiability conditions on the objective function). However, it was empirically proved in [10] that gradient descendant methods on formulation of Eq. 4 converge faster than cutting plane methods over formulation of Eq. 3.

2.2 Non-Sparse Lp-Norms

Sparse models like ℓ_1 norm formulations have a nice property of making solutions more interpretable due to the selection of only a few kernels from the list of all candidates. However, it has been demonstrated in [6] that sparse kernel mixtures do not always lead to best accuracy and therefore some others regularizers, like ℓ_2 norm, which lead to non-sparse kernel mixtures, might be better for improving accuracy.

$$\begin{aligned} &\min_{\{f_m\},b,\xi_i,d} \ \tfrac{1}{2}\sum_{m=1}^{M}\tfrac{1}{d_m}\|f_m\|^2_{\mathcal{H}_m}+C\sum_{i=1}^{n}\xi_i \\ &s.t.\forall_i\, y_i\left(\sum_{m=1}^{M}\langle f_m,\phi_m(x)\rangle_{\mathcal{H}_m}+b\right)\geq 1-\xi_i \\ &w.r.t. \quad d_m\in\mathbb{R}+,\ \ \|d\|_{\mathrm{p}}^{\mathrm{p}}\leq 1 \quad \xi_i\in\mathbb{R}_+ \\ &\qquad b\in\mathbb{R} \quad f_m\in\mathcal{H}_m \end{aligned} \tag{5}$$

More recently, [8] proposed a formulation where arbitrary ℓ_p norms induce a different degree of sparsity on the solutions (Eq. 5).

The above formulation is smooth, considering the same previously convened framework, and in order to extend the norm ℓ_1, the constraint $\|d\|_1 = 1$ is replaced by a $\|d\|_p = 1$ norm regularization, where *p* is an arbitrary value; however, convexity cannot be granted any longer, so the solution proposed by [8] consists in the relaxation of the constraint $\|d\|_p = 1$ to $\|d\|_p \leq 1$, which leads to Eq. 5 in order to apply convex optimization methods. In [8], it was proven that the optimal solution will always lay on the boundary of the search space where $\|d\|_p = 1$.

2.3 Proposed Numerical Schemes

In this work, we did enhance the MKL Newton's method in a wrapper approach with a modified Golden and Fibonacci linear search (see Algorithm 2) instead of using the common binary search available in Shogun toolbox and in [2]. Our second contribution is focused on the correction of the descent direction, for the cases when the Hessian matrix is singular, by updating the direction with the Reduced Gradient method. Our final contribution consists in an empirical demonstration that MKL models solved with the proposed algorithm outperform the Cutting Plane methods (in a wrapper approach), considering both accuracy and execution time. Formulations with different sparsity degrees can indeed lead to different accuracy results on Computer Vision recognition problems. We use a real recognition problem to illustrate this.

Several optimization models have been proposed in order to solve Eq. 4 and Eq. 5 such as wrapper approaches [2, 7, 8, 10] and interleaving approaches [6, 8, 12], among other approaches. In this paper, we will focus on wrapper-based approaches (see Algorithm 1), where d_m is optimized in the outer loop and SVM dual α coefficients are computed in the inner loop by using standard Quadratic Programming (QP) solvers.

Algorithm 1. General Scheme of MKL Wrapper-Based Approach Methods

Init $d_m = \frac{1}{M} \; \forall_{m=1}^{M}$
while conditions for d_m not met **do**
  update d_m with some specific method
  while conditions for α not met **do**
    solve α with SVM solver,
      where $K = \sum_{m=1}^{M} d_m K_m$
  end **while** inner loop
end **while** outer loop

Algorithm 2. Enhanced Newton Descendant in a Wrapper Approach for MKL

init $d_m = \frac{1}{M} \; \forall_{m=1}^{M}$
while (dual gap=0 | KKT) not met **do**
  compute ∇f_{obj} and Hessian F_{obj}
    for a fixed d_m
  if F_{obj} singular
   then compute $\theta = -F_{\text{obj}}^{-1} \nabla f_{\text{obj}}$
   else correct θ with ReducedGradient
  while f_{obj} decrease for a fixed γ
    update $d = d + \gamma\theta$
    solve α with SVM solver,
      where $K = \sum_{m=1}^{M} d_m K_m$
  end **while**
  Custom Golden/Fibonacci Search
    to find the best γ
  update $d = d + \gamma\theta$
end **while**

As mentioned above, d_m can be updated using analytical methods such as Newton's descent. Compared to other gradient methods, this method has a high convergence rate due to the inclusion of curvature information on the formulation by making use of the second order derivatives with the computation of the Hessian matrix. It is very suitable for large scale problems and a medium size list of kernels candidates; however, it requires the granting of the second order differentiability conditions for the objective

function. It is also known that the calculation of the gradient has a high computational cost and this issue is accentuated for techniques based on Newton's method, as it requires the calculation of the Hessian matrix; it is then recommended to use quasi-Newton's methods (i.e. Broyden-Fletcher-Goldfarb-Shanno, BFGS) for the case when the list of kernels candidates is large. Also, in the ℓ_p norm MKL formulation context, the global convergence of Newton's method cannot be granted due to the relaxation in Eq. 5 to enforce convexity. To diminish this global convergence issue, we propose Algorithm 2.

Algorithm 3. Cutting plane method in a wrapper approach for MKL

init $d_m = \frac{1}{M} \; \forall_{m=1}^{M}$

repeat

compute α with SVM solvers,

where $K = \sum_{m}^{M} d_m K_m$,

compute new cutting plane γ from α

include γ in active constraints set

update d_m with active constraints

until (gap < Epsilon) is met

After discussing pros and cons of the analytical update methods, we remark that the wrapper approach has also been used with the Cutting Plane method (see Algorithm 3), where d_m is updated in the inner loop. In this case, d_m is updated by solving a Linear Problem (LP) (for the ℓ_1 norm) or a Quadratic Constrained Linear Problem (QCLP) (for the ℓ_p norms, $1 < p < \infty$) with several and continuously added new generated constraints (see Algorithm 3). Although this method does not require the differentiability conditions on the formulation, its convergence rate has not been estimated and could reach a steady state for the objective function while still being far from convergence; also high computational costs are presented due to the update of d_m in every outer step with an LP or QP solver.

Some other combined methods have been proposed recently in order to improve the memory-less and slowly convergent cutting planes in a combined projection to the level set method converging much faster than any of the previous methods alone [3]. Also, recent works [13] formulate the ℓ_p norm MKL in a way that it is generalized to $p >= 1$, and the kernel coefficients are computed in a closed form very efficiently.

AU	Action	Description	Appearance
12	Oblique	Lip Corner Puller	

Fig. 1. Example of an action unit (AU)

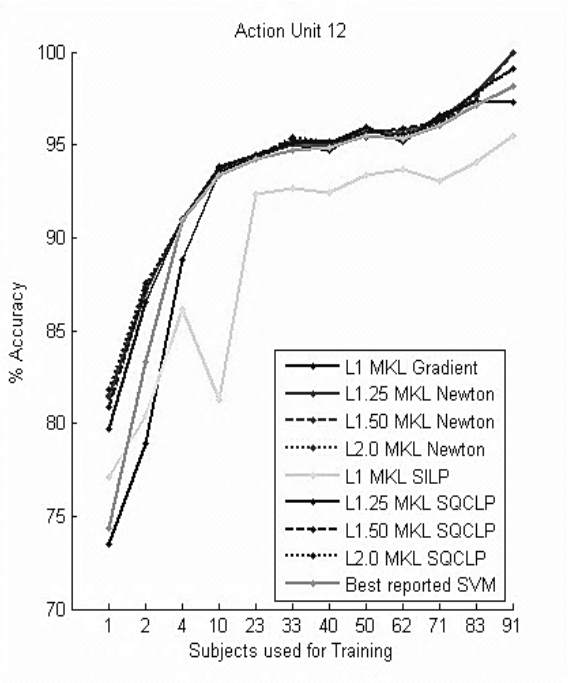

Fig. 2. Comparing accuracy values of models for different training sample sizes

3 Experimental Results

To illustrate the use of MKL in computer vision, we selected facial expression analysis as an application. To measure subtle facial changes, Ekman *et al*. [3] developed the Facial Action Coding System (FACS), which is a human-

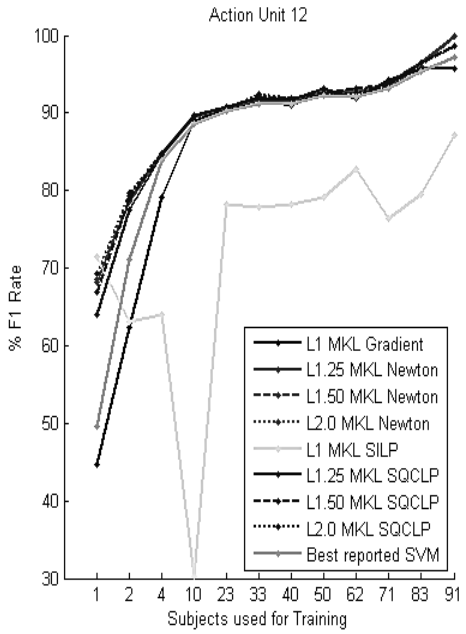

Fig. 3. Comparing F1 rate of models for different training sample sizes

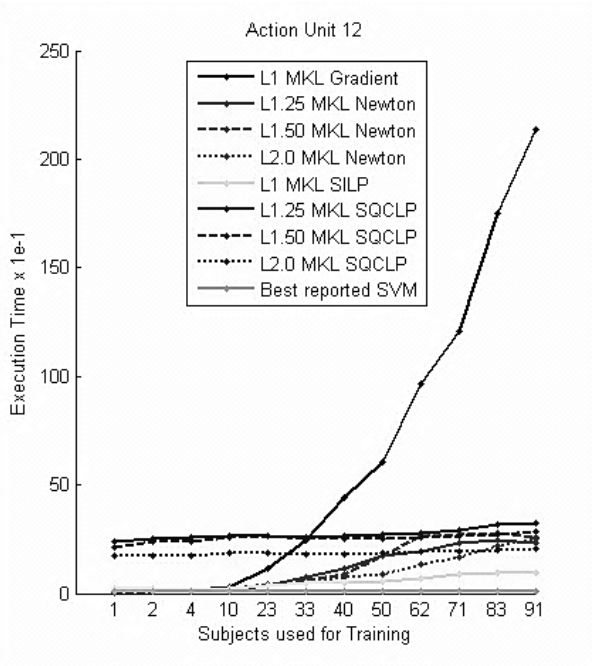

Fig. 4. Comparing execution time of models for different training sample sizes

observer-based system designed to detect subtle changes in facial features, and it describes facial expressions by action units (AUs). An example of an AU is given in Fig. 1.

To recognize AUs in image sequences, geometrical features [15], describing displacement of 83 facial feature points [4], are estimated. In this work, we compare our earlier findings for the recognition of lower AUs. In this case, only AU12 will be considered. The results are summarized in Table 1, and a comparison of accuracy, F1 rate, and execution time values is performed in Fig. 2, 3, 4 respectively.

The experimental settings are similar to the ones used in [10], with the difference that we use a list of 23 radial basis function (RBF) kernels as candidates on the MKL. The state-of-the-art solutions based on SVM models can be found in [4], the same SVM model has been used for comparison in Table 1.

We tested the chosen methods for different sizes of training data sets, selecting the samples randomly. This process is repeated 10 times in order to compute accuracy and F1 rates at the end. As it can be seen, the sparse ℓ_1 norm model, despite of not being the best model, is very competitive, but this is not surprising since the best reported model for the problem (in [4]) is

Table 1. Classification accuracy for different algorithms in different regularization norms: Best reported SVM (BSVM), Gradient (GD), Newton's (ND), SILP (S1), Sequentially QCLP(S2)

Training Samples		1	33	71	91
BSVM		74.37	94.73	96.01	98.18
GD	L1	73.49	95.03	**96.58**	97.27
ND	L1.25	81.41	94.99	96.36	**100.0**
ND	L1.50	81.41	95.21	96.01	**100.0**
ND	L2.00	81.52	95.29	96.01	99.09
S1	L1	77.10	92.64	93.05	95.45
S2	L1.25	79.68	95.12	96.47	97.27
S2	L1.50	80.91	95.21	96.24	99.09
S2	L2.00	**81.82**	**95.42**	96.13	99.09

within the list of candidate kernels and could be the choice for sparse models; however, it is reached and improved several times by other non-sparse models. For the AU12 recognition problems, it is recommended to use the $\ell_{1.25}$ norm, considering the data used in our experiment.

4 Conclusions

In this article, we have studied an application of different MKL ℓ_p norms formulations, optimization strategies, and algorithms for Computer Vision recognition problems, showing that MKL should be considered when facing such kind of problems. We have also shown that the sparsity/accuracy trade-off should be considered during the experimental process for each particular recognition problem as a part of the selection process of the regularizer of MKL formulation. Indeed, the regularizer highly depends on the correlation of features in the feature space and also on the training data nature. We have finally shown that our proposed algorithm is indeed more effective in terms of execution time and leads to a slightly better predictive performance.

References

1. **Bach, F., Lanckriet, G.R.G., & Jordan, M.I. (2004).** Multiple Kernel Learning, Conic Duality, and the SMO Algorithm. *Twenty-first international conference on Machine learning (ICML '04),* Banff, Alberta, Canada.
2. **Chapelle, O. & Rakotomamonjy, A. (2008).** Second order optimization of kernel parameters. *NIPS´08 Workshop: Kernel Learning - Automatic Selection of Optimal Kernels,* Whistler, Canada.
3. **Ekman, P., Friesen, W.V., & Hager, J.C. (2002).** *The Facial Action Coding System: The Manual on CD-ROM & Investigator's Guide (2nd Ed.).* Weston, Florida: Network Information Research Corporation.
4. **Gonzalez, I., Sahli, H., & Verhelst, W. (2010).** Automatic Recognition of Lower Facial Action Units. *7th International Conference on Methods and Techniques in Behavioral Research (MB'10),* Eindhoven, Netherlands, Article No.8.
5. **Hou, Y., Sahli, H., Ilse, R., Zhang, Y., & Zhao, R. (2007).** Robust Shape-based Head Tracking. *Advanced Concepts for Intelligent Vision Systems (ACIVS 2007), Lecture Notes in Computer Science*, 4678, 340–351.
6. **Kloft, M., Brefeld, U., Laskov, P., & Sonnenburg, S. (2008).** Non-sparse multiple kernel learning. *NIPS'08 Workshop: Kernel Learning - Automatic Selection of Optimal Kernels,* Whistler, Canada.
7. **Kloft, M., Brefeld, U., Sonnenburg, S., & Zien, A. (2010).** *Non-sparse regularization and efficient training with multiple kernels (Technical Report No. UCB/EECS-2010-21).* Berkeley: University of California at Berkeley.
8. **Kloft, M., Brefeld, U., Sonnenburg, S., Laskov, P., Müller K.R., & Zien, A.** (2009). Efficient and accurate Lp-norm multiple kernel learning. *Advances in Neural Information Processing Systems,* 22, 997–1005.
9. **Lanckriet, G.R.G., Cristianini, N., Bartlett, P., Ghaoui, L.E., & Jordan, M.I. (2004).** Learning the Kernel Matrix with Semidefinite Programming. *The Journal of Machine Learning Research,* 5(Jan), 27–72.
10. **Rakotomamonjy, A., Bach, F.R., Canu, S., & Grandvalet, Y. (2008).** SimpleMKL. *Journal of Machine Learning Research,* 9(Nov), 2491–2521.
11. **Schölkopf, B. & Smola, A.J. (2002).** *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Cambridge, Mass.: MIT Press.
12. **Sonnenburg, S., Rätsch, G., Schäfer, C., & Schölkopf, B. (2006).** Large Scale Multiple Kernel Learning. *Journal of Machine Learning Research,* 7(Jul), 1531–1565.
13. **Xu, Z., Jin, R., Yang, H., King, I., & Lyu, M.R. (2010).** Simple and Efficient Multiple Kernel Learning by Group Lasso. *27th International Conference on Machine Learning (ICML 2010),* Haifa, Israel, Paper ID: 540.

Mitchel Alioscha-Pérez received his B.Sc. (golden diploma) in Computer Science from the Central University of Las Villas (UCLV), Villa Clara, Cuba. He received his M.Sc. degree in Systems and Signals with an honorary mention in Digital Images and Signals Processing from UCLV, Cuba, in 2011. He is about to start his Ph.D. studies in topics of Computer Vision. His main research interests are image and video processing, machine learning, probabilistic models, and optimization methods.

Hichem Sahli received his Ph.D. degree in Computer Science from the *Ecole* Nationale Superior de Physique Strasbourg, Strasbourg, France, in 1991. He joined the Department of CAD and Robotics of the Ecole des Mine de Paris in 1992. In 1999, he moved to the Vrije Universiteit Brussel, Brussels, Belgium, where he is currently a professor at the Department of Electronics and Informatics. He coordinates a research team in Computer Vision and Audio Visual Signal Processing. His main research interests include computational vision, image and motion segmentation, machine learning, visual reconstruction and optimization. Dr. Sahli is a member of IEEE and ACM.

Isabel González received her M.Sc. degree in Computer Science from the Vrije Universiteit Brussel, Brussels, Belgium, in 2008. She is currently a Ph.D. candidate and a teaching assistant at the Department of Electronics and Informatics, Vrije Universiteit Brussel, Belgium. Her main research interest is in the field of facial expression analysis. She also collaborates in several research projects: the HOA Project CadE-games, Toward Cognitive Adaptive Edu-games, and the European Aliz-e project, Adaptives Strategies for Sustaintable Long-Term Social Interaction.

Alberto Taboada-Crispi graduated in Electronic Engineering from the Central University of Las Villas (UCLV), Villa Clara, Cuba, in 1985. He received his Master in Electronics from UCLV in 1997, and his Ph.D. from the University of New Brunswick, New Brunswick, Canada, in 2002. He worked as a Specialist for Medical and Laboratory Equipment in the Villa Clara Provincial Hospital from 1985 to 1988. Since then, he has been with the Faculty of Electrical Engineering of the UCLV. He is currently a Professor and a Researcher, and the Director of the Center for Studies on Electronics and Information Technologies (CEETI), UCLV. He is also a reviewer for Elsevier and Springer journals and a member of several Cuban science societies. His topics of interest include instrumentation, analog and digital processing of signals, digital processing of signals and images, and biomedical applications.

Article received on 02/02/2011; accepted on 05/10/2011.