Arm muscular effort estimation from images using Computer Vision and Machine Learning

Leandro Abraham[123] , Facundo Bromberg[13] and Raymundo Forradellas[2]

[1] Laboratorio DHARMa, DeSI, Universidad Tecnológica Nacional, FRM
leandro.abraham@frm.utn.edu.ar and fbromberg@frm.utn.edu.ar
[2] CEAL, Universidad Nacional de Cuyo, Facultad de Ingeniería
[3] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Abstract. A problem of great interest in disciplines like occupational medicine, ergonomics, and sports, is the measurement of biomechanical variables involved in human movement and balance such as internal muscle forces and joint torques. This problem is solved by a two-step process: data capturing using impractical, intrusive and expensive devices that is then used as input in complex models for obtaining the biomechanical variables of interest. In this work we present a first step towards capturing input data through a more automated, non-intrusive and economic process, specifically weight held by an arm subject to isometric contraction as a measure of muscular effort. We do so, by processing RGB images of the arm with computer vision (Local Binary Patterns and Color Histograms) and estimating the effort with machine learning algorithms (SVM and Random Forests). In the best case we obtained an FMeasure= 70.68%, an Accuracy= 71.66% and a mean absolute error in the predicted weights of 554.16 grs (over 3 possible levels of effort). Considering the difficulty of the problem , it is enlightening to achieve over random results indicating that, despite the simplicity of the approach, it is possible to extract meaningful information for the predictive task. Moreover, the simplicity of the approach suggests many lines of further improvements: on the image capturing side with other kind of images; on the feature extraction side with more sophisticated algorithms and features; and on the knowledge extraction side with more sophisticated learning algorithms.

Keywords: muscle arm effort, SVM , random forests, LBP , color histograms

1 Introduction

Musculoskeletal system biomechanics is a scientific discipline that aims to study the mechanical structures, laws, models and phenomenons that are important to the balance and movement of humans. The biomechanical variables most studied when analyzing balance and movement are internal and external muscles forces, and joint torques. The analysis of these variables allows the identification of harmful movements, overexertions, awkward postures, musculoskeletal disorders, optimal movements, among other states of the human body that have high impact in its health and performance. This results in its application in disciplines like occupational medicine [4], ergonomics [28], and sports [16], among others.

The estimation of internal muscular forces and joint torques is not made through direct measurement, but indirectly through dynamical models. These are: inverse dynamics, forward dynamics, and electromyography guided models [30,21,22]. Possible inputs for these models are the level of activation of the muscle, obtained by processing the electric signal (electric activity) produced by the muscles when contracting; the kinematic variables of joint positions at each instant; and the external forces involved in the movement or posture. After a computer process they return as output, among other variables, the internal forces, the joint torques or the muscular activation level. Nowadays, the measurement of electric muscular activity during muscle contractions is performed by an expensive device called electromyograph (EMG). The use of this device requires the adherence of wired electrodes to the skin or the introduction of wired needles in the muscles. Although there exists as well the wireless EMGs, these are considerably more expensive and therefore rarely used in practice. To capture the kinematic variables, commonly used devices are: goniometers to measure angles between body parts (require fixing sensors on the body) and motion capture systems that visually measure the positions of body parts (require adherence of markers and using expensive multi-camera systems), among others. The mentioned technologies are expensive, they limit the body movements through electrodes, needles, marker suites, and goniometers; and require special mounting devices; all of which makes them unsuitable for use outside a laboratory environment. In recent years, automatic measurement of kinematic information using inexpensive cameras has achieved a significant level of maturity because of the appearance on the market of low cost depth sensors (Microsoft Kinect [2], Asus Xtion [1], among others). With these devices it has been made possible to measure the joints position [10] with acceptable precision. However, to the best of the author's knowledge, there is still no convincing image-processing technology for estimating the level of muscular activation. This paper presents an approach that aims to take the first steps to solve this problem.

The objective of the authors research line is to solve the problem of estimating the electric muscular activity by the indirect estimation of highly correlated variables, in particular for this paper, the weight of objects held in a static posture of isometric contraction. The correlation is justified by considering that the greater the weight, the greater will be the force needed to hold it, and greater the electrical activity produced to contract the muscle. We refer to these measures now on under the name of *muscular effort* or *muscular activity*. This line of research has practical applications in situations where it is necessary to estimate the muscular effort wirelessly. For example, it could be used to estimate the muscular activation patterns of an athlete during the execution of some movement. Another application could be to estimate the effort that a worker is performing during a task in order to detect harmful movements in an occupational environment. It's worth to clarify that these applications will be possible when this research line has reached maturity and when the framework has been tested in real conditions. This work presents the first experimentation for the research line, lacking of a practical application in the immediate future.

The main contribution of the present work is the application of computer vision and supervised machine learning for solving the problem of estimating the weight of object held by the arm of humans through postures of isometric contraction, from RGB images of the arm taken in uniform scenes, with the same conditions of illumination, scale and point of view for each one.

For shading some light on the originality of our contributions, its impact, and feasibility we present related works in Section 2. We present our approach for solving the problem in Section 3. Section 4 presents the experimental setup used for proving the effectiveness of our approach, followed by its results in Section 5. We end with discussions in Section 6, and conclusions in Section 7.

2 Related work

As far as the authors have been able to investigate, there are no contributions in the literature that solve the specific problem of estimating muscular effort, exclusively from skin images. However, it is possible to find work related to the general problem of characterizing muscle contraction from body images but these either use extra information such as kinematic variables obtained through other capturing devices or EMG measurements, or rather aims to model the skin features for different underneath contractions. In this section we compare our contribution with the problems and techniques considered by these works, highlighting the differences that justify the originality of our proposal.

We start by the paper [5], whose main contribution is the interactive presentation of the approximated level of muscular activation that is produced in inferior extremities muscles when executing different movements. This information is presented in an augmented reality interface and the level is obtained from a database generated from EMG measurement and its respective kinematic information. This database is a straightforward indexing mapping between kinematic information and EMG measurements, with no model learning that maps them, and therefore prone to errors when used for estimation of muscle activity. When compared to our approach, it is based on kinematic information obtained by skeleton tracking from an RGBD sensor, and not on skin RGB images.

Another related problem is the one discussed by [26], consisting in the identification and classification of walking patterns between examples of healthy and injured subjects. Here, they use skin deformation information, together with EMG measurements of the subjects, as input to a neural network for the classification. Although this paper focuses on computer vision for extracting skin deformation information from RGB images, they don't use this skin information for estimating muscular activity, but rather use it with such information, obtained by direct measurements of EMGs.

Another group of works have been dedicated to solve the problem of modeling the deformation of the skin due to underlying muscle activity [23,18,24,12] [34,11,29], but with applications in computer graphics for animating virtual characters. The problem solved by these papers differ from the one that we plan to solve in that they don't attempt to estimate the muscular activity, but rather seeks to model the deformation due to muscle contraction, without prediction.

Another difference is that they obtain the visual information from motion capture systems by marker tracking and depth information from RGBD sensors, instead of from RGB images as it is done in this work.

There is also a group that have used computer vision techniques to evaluate and measure the deformation and movement produced in the skin due to muscular contraction mainly using pixel movement information form RGB images [6,17,35,9]. The latest work of Carli et. al, as asserted by the authors, is an improvement of the former, and both [6,17] aims at modeling skin deformation when the biceps is under isometric contraction. Although clearly related to our present work, their goal is on the opposite direction, being our goal the modeling of the underneath contraction from skin features, and not the other way around. It is worth to clarify that [35,9] work over an insect instead of over human skin.

Finally there is a group of works that pretend to solve the problem of prediction of electric muscular activity from kinematic information, most of them validating with EMG captures [31,8,13,25,32,3] or by muscular activation obtained from an inverse dynamic model [15]. Their approaches consider methods as neural networks [15,3,25,8], probabilistic Bayesian reasoning [3,31], curve fitting methods [3], regression models [32] and biomechanical models [13]. While these works solve the same problem we want to solve, they do so with kinematic information obtained from motion capture systems instead of visual skin information obtained from RGB images.

From the survey of the state of the art, we can argue that there is no previous work that performs estimation of muscular effort from external images of the skin using Computer Vision and Machine Learning, as we propose in this paper.

3 Our approach

Our approach is an autonomous modeling of the relationship between skin characteristics and weights lifted by an arm in conditions of isometric contraction. This modeling is constructed by a supervised learning procedure outlined in Figure 1. This general approach can be instantiated differently for the different components and we discuss these variations in this section. The input of the procedure are images of the arm region, corresponding to different conditions of weight lifting. Input images are cropped to the biceps region and segmented from the blue background ((A), detailed in sub-section 3.1). By a feature extraction process, each image is characterized in a vector of numeric values ((B), detailed in sub-section 3.2). Optionally these vectors could be complemented with anthropometric data of the subject ((C), detailed in sub-section 3.3). These vectors could be also normalized in relation to minimum effort cases for the corresponding subject ((D), detailed in sub-section 3.4). The resulting vectors are labeled with the corresponding weights and supplied as input to a machine learning algorithm during the learning stage ((E), detailed in sub-section 3.5).

3.1 Capture and segmentation (A)

The process starts by capturing static RGB images, with blue background, of subjects arms holding one of five possible objects of differing weights. Then, the images are cropped to the region of the biceps. After that, the blue background

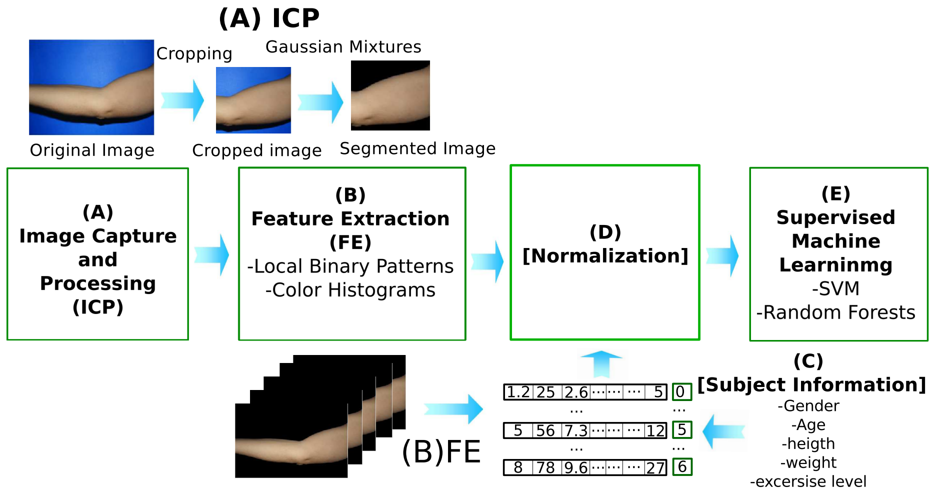

Fig. 1. Outline of the approach

is removed by automatic segmentation process using the Gaussian Mixtures algorithm [33], pre-trained with some manually segmented images.

3.2 Feature extraction (B)

Each segmented and cropped image is processed for feature extraction producing a vector of a fixed number of quantities (i.e., dimensions) that describes the image. The feature vector of training images is extended with a label indicating the actual weight sustained. In this work we evaluated our approach building the feature vector with information of texture and/or color.

For texture we used the Local Binary Patterns [19] feature extraction algorithm. Briefly, it proceeds by dividing the image into 9 regions of interest (ROI) through a 3 by 3 grid. Then, local binary patterns are computed for each of these ROI, generating for each of its pixels a binary code of 8 bits, one for each of its 8 nearest neighbors, deciding 0 if the pixel is greater than its neighbor or 1 otherwise. For each ROI the binary codes are accumulated in a histogram according to their decimal value, and these histograms are later concatenated in a vector that describes the whole image. We used 59 bins for each ROI histogram.

For color we used color histograms [20,27]. As in the previous case, images were divided into 9 ROIs through a 3 by 3 grid, and for each of these a color histogram was computed using the HS channels of the HSV color space. We used 5 bins to accumulate the colors of the pixels of each channel. The image was then described through a concatenation of the color histograms of each of the ROIs.

3.3 Inclusion of anthropometric data (C)

Another component of the approach is the possibility to augment the raw feature vector obtained by image description with anthropometric information of the subject. The anthropometric data included in this step is: the gender of the subject, his weight, his height, his body mass index, the level of physical exercise that he normally does, and the level of exercise that he normally does with the arm (all of these obtained by questioning the subject during the picture session).

3.4 Feature vector normalization (D)

It is quite clear that generalizing among subjects is a difficult task, as their arms may differ in so many aspects that has no relation whatsoever with the muscle contraction, adding a lot of noise. Moreover, their arms may produce quite different skin patterns for the same muscle contraction. We propose here a first step to tackle these issues that could be thought as a calibration process, where the feature vector of some subject is normalized by subtracting from it the feature vector of the case of effort zero. This procedure breaks the standard supervised learning protocol, as for new subjects as input to the framework, it is provided not only the input data, but the output data of the zero effort cases. In practice, however this is a simple request for the subjects.

3.5 Estimation of muscular effort by machine learning (E)

The final component of our approach consists in running a supervised learning algorithm for predicting the weight lifted by a biceps muscle, when given images of the arm. In this work we evaluate the performance of radial basis Support Vector Machine [7] (SVM) and Random Forests [14]. In a first stage the algorithms are trained using a corpus of labeled subjects images through a dataset whose data-points consists of the feature vector produced as explained above, that may be normalized or not and may contain anthropometric data or not, and labeled with the actual weight lifted in that image.

During training we tuned the user-given parameters of the learning algorithms over a grid of possible values using a 10 fold cross-validation over the training set, choosing the parameters with overall better performance over the 10 cases. Radial SVM has two given parameters: C (the soft margin), and σ (the standard deviation of the kernel parameter), both tuned over the set of values $\{0.002, 0.02, 2.20, 200\}$. Random Forests was tuned over the number of random variables used for constructing each of its trees, considering different proportions over the total number of variables (of the input feature vectors): $\{2\%, 6\%, 12\%, 25\%, 50\%, 75\%, 100\%\}$. At the moment of the estimation of the effort level of a new image, the trained model is used.

As there have not been found previous work that estimate the muscular effort from images using a similar approach, we decided to include a third learning algorithm, the random classifier, as our competitor.

4 Experimental Setup

The current and following sections presents our experimental setup. With these experiments we intend to answer empirically the main question of this work: whether computer vision and supervised-learning techniques are capable of extracting enough information from arm images for inferring the weight they are lifting. For that, we start presenting in the next sub-section 4.1 the process followed for capturing the image corpus, followed by sub-section 4.2 that explains the process followed for generating the training and testing datasets, and conclude with the sub-section 4.3 that describes the performance measures used for evaluating the outputs of the machine learning algorithms.

4.1 Capture of Raw Image Data

For this work we captured images from 100 human subjects. The group of subjects presents variability in their characteristics. It is conformed by both women and men of white to latin dark skin that exercise with differing levels of intensity from low to moderate, with ages ranging between 17 and 54 years old, weights ranging between 47 and 113 kgs, their heights varying between 1.6 mts to 1.9 mts, and their body mass index varying from 19.32 and 34.16. For each subject, 3 photos holding 5 objects of different weights were captured using a Canon EOS 1000D camera with Zoom Lens of 75-300mm, without flash and manual focus. The same set of objects were used for each subject: object $O0$ of 0 grs, the object $O500$ of 500 grs, the object $O1k$ of 1300 grs, the object $O2k$ of 1900 grs, and the object $O4k$ of 3800 grs. The photos were captured as close as possible of the biceps, using a blue background (to ease the autonomous segmentation of the biceps). Each subject was told to exert an isometric contraction effort, holding in front of the camera a posture where the segment of the arm from the shoulder to the elbow was separated approximately 45°from the torso and the segment of the arm from the elbow to the hand was horizontal. All the photos were captured over this posture in order to control this variability for the experiments. The scene was artificially illuminated with a little searchlight of 60W, located in front of the subject, but from the right side, elevated about 1.5 meter from the subject, and aiming to the arm. As mentioned earlier, each image was then cropped to the biceps zone in order to remove clutter from the image, and then segmented from the background.

4.2 Datasets generation

Our approach is tested in the standard training/testing approach, where the model is learned over the training set, and evaluated over the test set. We generated 10 pairs of training/testing datasets, each pair generated by first selecting randomly 8 out of the 100 subjects (8%), and producing as testing set the 120 images corresponding to 8 subjects times 5 weights times 3 (repetitions) images; and as training set the images of the remaining 92 subjects (92%). The experimental results shown in the following section are over different instances of the approach, and for each case it is reported the *mean* and *standard deviation* over the 10 datasets pairs.

4.3 Performance measures

In this section we present the performance measures used for the evaluation of each instance of the proposed framework. As the problem is solved as one of image classification, we report the classical performance measures for classification tasks: accuracy over all the confusion matrix, named as Overall Accuracy and computed as OvA =sum of the diagonal values / sum of the whole matrix values, also Precision (P), Recall (R), FMeasure (FM), and Accuracy (A); computed for each class independently, resulting in n of them (n = number of classes), and grouped by reporting their means over all the classes. For each class i, they are computed as follows: P_i = number of correctly classified examples of class i / total number of examples classified as class i; R_i =number of correctly classified objects of class i / total of examples that truly are of class i, $FM_i = 2 \times ((P_i \times R_i)/(P_i + R_i))$, and finally A_i =sum of the quantities of the diagonal + the sum of the quantities not belonging neither to the row i nor to the column i / total number of examples. We also report the *mean absolute error* (MAE) by transforming the object class label to its corresponding value in grams. This measure tells us how is the error made at estimating the weight of the held out object. It is calculated as the mean of the absolute values of the differences between the predicted weight minus the true weight. Therefore, the smaller the MAE, the better the result of the estimation.

5 Experimental Results

To conclude the validation of our hypothesis and the contribution of this paper, we present in this section the results of the experiments over all instantiations resulting from selecting the learning algorithm, the feature extraction algorithm, whether it is used anthropometric data or not, and whether a normalization is conducted over the dataset or not. Each case is ran over all 10 training datasets, and the mean and standard deviation (subindex and parenthesis numbers) of the performance measures over the corresponding test sets are reported. The results are shown in Tables 1 (classification over 3 levels) and 2 (classification over 5 levels). The first column labeled 'Desc' describes the feature extraction technique used, with LBP for Local Binary Patterns and HSCH for Color Histograms. The second column labeled 'SI' indicates whether anthropometric subject information is appended to the input dataset set or not. The third column labeled 'N' indicates whether normalization was applied or not. The fourth column 'MLAlg' describes the machine learning algorithm used, with SVM for Support Vector Machine, RF for Random Forests and RDM for the Random classifier. The remaining columns report each of the 6 performance measures.

We start with a simpler setup where we tried to predict only over 3 levels of effort, considering only the examples of $O0$, $O2k$ and $O4k$. These results are shown in Table 1. As it can be seen in the Table, the prediction made with the proposed descriptors when normalization is not performed (not grayed out) were better than the random prediction. Despite of this, they present low classification measures and relatively high MAE. In those cases where normalization was applied (grayed out), the results became substantially better. In the best cases we obtain classification measures over 70% and a MAE down to 554 grs. We also tested the performance of an instance of the approach that use as descriptor the combination of the two descriptors (HSCH&LBP), showing only minor improvements over the overall best cases, but large improvements for the SVM classifier. With this first experiment it is possible to prove that the normalization of the raw feature vectors proposed improves the performance of the framework, leading to acceptable values of classification measures and weight prediction.

In Table 2 we show the results of a second experiment. In order to improve the impact of our solution, we decided to evaluate the approach in the task of classification over the 5 levels of effort. As it is possible to see, when in-

Desc.	SI	N	MLAlg.	$\hat{P}$	$\hat{R}$	$\hat{FM}$	$\hat{A}$	OvA	MAE
RDM				*34.7* (5.81)	*34.72* (6.03)	*34.36* (5.92)	*56.48* (4,02)	*34.72* (6.03)	*1657.22* (147.08)
LBP	NO	NO	RF	49.38 (9.85)	49.58 (7.52)	48.39 (8.2)	66.38 (5)	49.58 (7.52)	1095.13 (219.46)
LBP	YES	NO	RF	49.3 (7.32)	49.44 (6.52)	48.48 (6.65)	66.29 (4.34)	49.44 (6.52)	1137.36 (182.17)
LBP	YES	YES	RF	71 (6.43)	69.86 (5.36)	69.77 (5.32)	79.9 (3.57)	69.86 (5.36)	572.63 (101.86)
LBP	NO	YES	RF	71.21 (7.65)	70.69 (7.24)	70.55 (7.22)	80.46 (4.83)	70.69 (7.24)	556.8 (137.65)
LBP	NO	NO	SVM	46.87 (8.02)	43.61 (6.58)	40.32 (7.54)	62.4 (4.39)	43.61 (6.58)	1369.58 (216.69)
LBP	YES	NO	SVM	48.74 (12.77)	48.05 (11.86)	45.36 (14.05)	65.37 (7.9)	48.05 (11.86)	1184.86 (379.45)
LBP	NO	YES	SVM	67.96 (5.75)	68.61 (4.73)	65.91 (6.74)	79.07 (3.15)	68.61 (4.73)	606.94 (97.95)
LBP	YES	YES	SVM	69.13 (13.56)	66.94 (3.57)	60.06 (4.87)	77.96 (2.38)	66.94 (3.57)	641.25 (72.58)
HSCH	NO	NO	RF	48.02 (9.87)	48.47 (8.62)	47.27 (9.17)	65.64 (5.74)	48.47 (8.62)	1169.02 (265.65)
HSCH	YES	NO	RF	53.32 (10.19)	53.33 (8.61)	52.06 (9.07)	68.88 (5.74)	53.33 (8.61)	1081.94 (207.04)
HSCH	YES	YES	RF	71.18 (6.29)	**71.66** (5.82)	70.4 (6.37)	**81.11** (3.88)	**71.66** (5.82)	572.63 (136.86)
HSCH	NO	YES	RF	70.31 (4.68)	69.72 (4.57)	69.14 (4.46)	79.81 (3.04)	69.72 (4.57)	606.94 (86.18)
HSCH	NO	NO	SVM	47.05 (8.35)	45.97 (8.11)	44.02 (8.55)	63.98 (5.4)	45.97 (8.11)	1377.5 (284.97)
HSCH	YES	NO	SVM	48.34 (9.65)	46.38 (6.9)	44.22 (7.43)	64.25 (4.6)	46.38 (6.9)	1385.41 (172.03)
HSCH	NO	YES	SVM	67.24 (6.62)	66.11 (5.32)	64.02 (6.07)	77.4 (3.55)	66.11 (5.32)	691.38 (106.14)
HSCH&LBP	NO	YES	RF	69.52 (5.12)	68.47 (5.03)	68.12 (4.73)	78.98 (3.35)	68.47 (5.03)	599.02 (95.59)
HSCH&LBP	YES	YES	RF	71.3 (5.46)	70.83 (5.27)	**70.68** (5.21)	80.55 (3.51)	70.83 (5.27)	**554.16** (100.29)
HSCH&LBP	NO	YES	SVM	**71.74** (7.98)	71.52 (7.11)	70.35 (7.84)	81.01 (4.74)	71.52 (7.11)	556.8 (149.51)

Table 1. Estimation results for 3 levels of effort. Cases with normalization were grayed out. Best performance measure obtained results are in bold.

creasing the number of classes, the performance measures decreased. This happens because five different levels of effort are less discriminable visually than three. Here we can see also that the normalization step improves significantly the performance of the framework. Also in this case, the combined descriptors got good results but they aren't significantly better than those obtained by the single descriptors. In these experiments also the case that got the best measures (HSCH&LBP+NO+YES+SVM) shows that combining the two descriptors make possible to get better results for the SVM classifier. It is possible to see that the classification measures obtained with normalization almost duplicate the values of the random classifier. Besides, in the better case, the MAE decreases significantly leading to an error of 825 grs which to the authors criterion is a good value considering that the minimum difference between two different examples could be 500 grs and the maximum difference could be up to 3800 grs.

Desc.	SI	N	MLAlg.	$\hat{P}$	$\hat{R}$	$\hat{FM}$	$\hat{A}$	OvA	MAE
RDM				*20.7* (3.76)	*20.41* (3.17)	*20.39* (3.39)	*68.16* (1.26)	*20.41* (3.17)	*1413.66* (95.94)
LBP	NO	NO	RF	34.62 (7.34)	32.51 (5.81)	31.61 (6.03)	73.01 (2.33)	32.52 (5.82)	1122.72 (195.39)
LBP	YES	NO	RF	26.12 (5.71)	27.25 (4.23)	25.51 (4.85)	70.9 (1.69)	27.25 (4.23)	1176.08 (150.4)
LBP	YES	YES	RF	45.48 (6.84)	46.33 (5.48)	45.14 (5.82)	78.53 (2.19)	46.33 (5.48)	826.16 (130.32)
LBP	NO	YES	RF	47.38 (5.93)	46.52 (5.9)	45.54 (5.75)	78.61 (2.35)	46.53 (5.89)	878.88 (143.88)
LBP	NO	NO	SVM	33.44 (9.13)	28.83 (3.89)	25.32 (3.45)	71.53 (1.55)	28.83 (3.89)	1499.75 (203.72)
LBP	YES	NO	SVM	26.89 (7.03)	28.01 (3.75)	24.96 (4.4)	71.21 (1.49)	28.04 (3.74)	1388.68 (190.97)
LBP	NO	YES	SVM	45.31 (5.23)	46.73 (4.63)	44.09 (5.17)	78.72 (1.84)	46.81 (4.62)	825.05 (212.92)
LBP	YES	YES	SVM	42.97(11.3)	43.33 (4.23)	36.17 (5.52)	77.33 (1.69)	43.33 (4.23)	1267.75 (123.59)
HSCH	NO	NO	RF	30.91 (4.86)	30.76 (5.1)	29.1 (4.69)	72.31 (2.04)	30.77 (5.11)	1068.86 (120.98)
HSCH	YES	NO	RF	29.58 (5.36)	29.91 (4.31)	28.87 (4.95)	71.96 (1.72)	29.91 (4.31)	1098.75 (129.61)
HSCH	YES	YES	RF	45.6 (4.45)	46.61 (5.29)	44.94 (4.73)	78.65 (2.12)	46.62 (5.3)	**825.43** (109.69)
HSCH	NO	YES	RF	46.52 (5.77)	47.44 (4.1)	45.13 (4.98)	78.98 (1.64)	47.95 (4.11)	890.3 (103.48)
HSCH	NO	NO	SVM	26.25 (6.54)	27.08 (5.73)	25.24 (5.77)	70.83 (2.29)	27.08 (5.73)	1380.58 (198.27)
HSCH	YES	NO	SVM	23 (3.81)	25.58 (3.49)	22.64 (3.42)	70.23 (1.39)	25.58 (3.49)	1418.75 (136.12)
HSCH	NO	YES	SVM	45.21 (3.37)	45.41 (2.29)	41.79 (2.39)	78.16 (0.91)	45.41 (2.29)	1076.41 (65.72)
HSCH&LBP	NO	YES	RF	45.32 (5.59)	45.66 (5.39)	44.23 (5.13)	78.26 (2.15)	45.66 (5.39)	857.83 (107.03)
HSCH&LBP	YES	YES	RF	45.54 (5.15)	46.1 (6.38)	45.01 (5.55)	78.44 (2.54)	46.11 (6.36)	824.43 (151.76)
HSCH&LBP	NO	YES	SVM	**51.67** (5.47)	**51** (4.09)	**48.33** (4.79)	**80.4** (1.63)	**51** (4.09)	842.66 (151.74)

Table 2. Estimation results for 5 levels of effort. Cases with normalization were grayed out. Best performance measure obtained results are in bold.

From the results we can affirm that it is possible to differentiate the level of muscular effort that a person is performing from a photo of a biceps: between 5 levels with FMeasure$\approx$ 48% and Accuracy$\approx$ 80%; and between 3 levels with FMeasure$\approx$ 70% and Accuracy$\approx$ 81%. Besides, it is possible to estimate the weight that a person is holding with his arm from a photo of his biceps: over 5 different weights (0 grs, 500 grs, 1300 grs, 1900 grs and 3800 grs) with a mean absolute error$\approx$ 825 grs; and over 3 different weights (0 grs, 1900 grs and 3800 grs) with a mean absolute error$\approx$ 554 grs.

6 Discussions

From the experimental results analysis showing improvements over the random classifier, we can conclude that our approach performs acceptably in finding visual patterns of biceps images and their correlations with muscle effort. This is clearly an important and solid step for achieving in the future a practical application. Despite this positive result, the framework of feature extraction followed by supervised learning presented in this work is proven by our experiments insufficient for successfully estimating muscular effort with enough precision for practical use for several limitations. One that we believe is the most limiting factor is the amount of training samples that, although considerably large, is still insufficient when compared with the great variability among the human subjects, resulting in an under represented underlying distribution. Another limitation is the simplistic approach considered, specially the use, over the shelf, of general purpose feature extraction algorithms that uses only color information to infer the latent, more informative 3D contour information. The practicality of the approach is also limited for the cases that uses the normalization step, where the prediction stage over unseen subjects requires labeled images for the case of zero effort, as well as by the requirement of exposed skin images on the area to be sensed. In conclusion, our approach shows an important and solid step toward a practical application in the future, and through the limitations described, it helps in highlighting the best possible future steps to follow.

7 Conclusions and future work

In this work we present the first steps towards estimation of muscular activity from skin images using Computer Vision and Machine Learning. The proposed approach consists in a framework that considers the capture of images of the area of the muscle, the generation of a raw feature vector for the images using image descriptors, the optional inclusion of anthropometric subject information to the raw feature vector, an also optional step of normalization of the feature vector, and finally the training of a model and estimating of level of new examples using supervised machine learning techniques. The first contribution of this work is the experimentation of this framework over an image dataset of 100 subjects performing 5 and 3 different levels of effort in static posture of isometric contraction, being this the first steps in this line of research. In this work we propose to use HS Color Histograms and Local Binary Patterns as image descriptors. Regarding the machine learning techniques used in this work, they are Support Vector Machines and Random Forests. Another contribution is the proposal of feature vector normalization in relation to the feature vectors of minimum effort, in order to improve the estimation results helping the framework to better generalize over previously unseen subjects.

As future work we plan to try other ways of improving the performance of the approach. As the exploration of description techniques and parameters of the machine learning algorithms has been very limited, we will extend the experimentation to other image description techniques to build the feature vector, and to other parameters for the machine learning algorithms. Another alternative is to use deep learning, a technology that combines feature extraction and model learning in one framework. As it is showed in this work, calibration is required to get good results, so we will design and experiment other ways of calibration. We also will extend the dataset with more subjects in order to find tendencies of improvement of the results as the number of train subjects increases.

Acknowledgements We thank CONICET for the funding of the author Leandro Abraham through a doctoral grant under the supervision of Dr. Raymundo Forradellas and the thesis supervision of Dr. Facundo Bromberg.

References

1. Asus xtion. `http://www.asus.com/latin/Multimedia/Xtion/`
2. Microsoft kinect sensor. `http://www.xbox.com/es-ES/Kinect`
3. A, J.L., J, F.A.: Evaluation of probabilistic methods to predict muscle activity: implications for neuroprosthetics. J of neural engineering 6(5), 055008 (2009)
4. B, C.D., Gunnar, A., J, M.B., et al.: Occupational biomechanics. Wiley New York (1984)
5. Bauer, A., Paclet, F., Cahouet, V., Dicko, A.H., Palombi, O., Faure, F., Troccaz, J., et al.: Interactive visualization of muscle activity during limb movements: Towards enhanced anatomy learning. In: Eurographics Workshop on VCBM (2014)
6. Carli, M., Goffredo, M., Schmid, M., Neri, A.: Study of muscular deformation based on surface slope estimation. In: Electronic Imaging. pp. 60640U–60640U. International Society for Optics and Photonics (2006)
7. Cortes, C., Vapnik, V.: Support-vector networks. Mach Learn 20(3), 273–297 (1995)
8. D, M.R., A, A.H., R, J.R., Otman, B.: A neural network model for reconstructing emg signals from eight shoulder muscles: Consequences for rehabilitation robotics and biofeedback. J of biomechanics 39(10), 1924–1932 (2006)
9. Davide, Z., Andrea, G., Vincent, T.: The use of optical flow to characterize muscle contraction. J of neuroscience methods 110(1), 65–80 (2001)
10. Dutta, T.: Evaluation of the kinect$^{\mathrm{TM}}$ sensor for 3-d kinematic measurement in the workplace. Applied ergonomics 43(4), 645–649 (2012)
11. Il, P.S., K, H.J.: Data-driven modeling of skin and muscle deformation. ACM TOG 27(3), 96 (2008)
12. L, A., M, B., N, M.T.: Dynamic skin deformation based on biomechanical modeling. In: Proc. of the 25th CASA. vol. 2 (2012)
13. Laursen, B., Søgaard, K., Sjøgaard, G.: Biomechanical model predicting electromyographic activity in three shoulder muscles from 3d kinematics and external forces during cleaning work. Clinical Biomechanics 18(4), 287–295 (2003)
14. Leo, B.: Random forests. Machine learning 45(1), 5–32 (2001)

15. Marzieh, M., Ling, W., Qin, L., Yaxiong, L., Jiankang, H., Dichen, L., Zhongmin, J.: Muscle activity prediction using wavelet neural network. In: ICWAPR 2013. pp. 241–246. IEEE (2013)
16. McGinnis, P.: Biomechanics of sport and exercise. Human Kinetics (2013)
17. Michela, G., Marco, C., Silvia, C., Daniele, B., Alessandro, N., Tommaso, D.: Evaluation of skin and muscular deformations in a non-rigid motion analysis. In: Medical Imaging. pp. 535–541. International Society for Optics and Photonics (2005)
18. Neumann, T., Varanasi, K., Hasler, N., Wacker, M., Magnor, M., Theobalt, C.: Capture and statistical modeling of arm-muscle deformations. In: Computer Graphics Forum. vol. 32, pp. 285–294. Wiley Online Library (2013)
19. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recogn 29(1), 51–59 (1996)
20. Olivier, C., Patrick, H., N, V.V.: Support vector machines for histogram-based image classification. Neural Networks, IEEE Transactions on 10(5), 1055–1064 (1999)
21. Pandy, M.G., Barr, R.E.: Biomechanics of the musculoskeletal system. Standard Handbook of Biomedical Engineering and Design (2004)
22. Qi, S., Buchanan, T.S.: Electromyography as a tool to estimate muscle forces. Standard handbook of biomedical engineering & design (2004)
23. Robertini, N., Neumann, T., Varanasi, K., Theobalt, C.: Capture of arm-muscle deformations using a depth-camera. In: Proceedings of the 10th European Conference on Visual Media Production. p. 12. ACM (2013)
24. Sand, P., McMillan, L., Popović, J.: Continuous capture of skin deformation. ACM Transactions on Graphics (TOG) 22(3), 578–586 (2003)
25. SD, P., AE, P., DA, S.: Artificial neural network model for the generation of muscle activation patterns for human locomotion. J of electromyography and kinesiology 11(1), 19–30 (2001)
26. Senanayake, S., Triloka, J., Malik, O.A., Iskandar, M.: Artificial neural network based gait patterns identification using neuromuscular signals and soft tissue deformation analysis of lower limbs muscles. In: IJCNN. pp. 3503–3510. IEEE (2014)
27. Shamik, S., Gang, Q., Sakti, P.: Segmentation and histogram generation using the hsv color space for image retrieval. In: ICIP. vol. 2, pp. II–589. IEEE (2002)
28. Shrawan, K.: Biomechanics in ergonomics. CRC Press (1999)
29. Sifakis, E., Neverov, I., Fedkiw, R.: Automatic determination of facial muscle activations from sparse motion capture marker data. In: ACM TOG. vol. 24, pp. 417–425. ACM (2005)
30. T, M.K., S, B.T.: Biomechanics of human movement. Standard handbook of biomedical engineering & design p. 26 (2004)
31. V, A.C., J, F.A.: Probability-based prediction of activity in multiple arm muscles: implications for functional electrical stimulation. J Neurophysiol 100(1), 482–494 (2008)
32. Xu, X., W, M.R., Jia-Hua, L.: A regression model predicting isometric shoulder muscle activities from arm postures and shoulder joint moments. J of Electromyography and Kinesiology 24(3), 419–429 (2014)
33. Yang, M.H., Ahuja, N.: Gaussian mixture model for human skin color and its applications in image and video databases. In: Electronic Imaging'99. pp. 458–466. International Society for Optics and Photonics (1998)
34. Youn, H.Q., Il, P.S., K, H.J.: A data-driven segmentation for the shoulder complex. In: Computer Graphics Forum. vol. 29, pp. 537–544. Wiley Online Library (2010)
35. Zoccolan, D., Torre, V.: Using optical flow to characterize sensory-motor interactions in a segment of the medicinal leech. J Neurosci 22(6), 2283–2298 (2002)