Visual interaction including biometrics information for a socially assistive robotics platform

Pierluigi Carcagnì, Dario Cazzato, Marco Del Coco, Cosimo Distante, and Marco Leo

National Research Council of Italy - Institute of Optics, Arnesano (LE), Italy

Abstract. This work introduces biometrics as a way to improve human-robot interaction. In particular, gender and age estimation algorithms are used to provide awareness of the user biometrics to a humanoid robot (Aldebaran NAO), in order to properly react with a specific gender/age behavior. The system can also manage multiple persons at the same time, recognizing the age and gender of each participant. All the estimation algorithms employed have been validated through a k-fold test and successively practically tested in a real human-robot interaction environment, allowing for a better natural interaction. Our system is able to work at a frame rate of 13 fps with 640×480 images taken from NAO's embedded camera. The proposed application is well-suited for all assisted environments that consider the presence of a socially assistive robots like therapy with disable people, dementia, post-stroke rehabilitation, Alzheimer disease or autism.

1 Introduction

The human-robot interaction (HRI) for socially assistive robotics (SAR) applications is a new, growing, and increasingly popular research area at the intersection of a number of fields, including robotics, computer vision, medicine, psychology, ethology, neuroscience, and cognitive sciences. New applications for robots in health and education are being developed for broad populations of users [37]. In this field, the introduction of biometrics can substantially increase the level of realism perceived by potential users, since they can be used in order to give to the person the opportunity to interact with an entity that can change its behavior depending on observed peculiarities. In particular, in [21] there is a definition of *soft biometrics* as characteristics that provide some information about the individual, but such that they lack of the distinctiveness and permanence to sufficiently differentiate any two individuals. The soft biometric traits can either be continuous (e.g., height and weight) or discrete (e.g., gender, eye color, ethnicity, etc.).

Beyond realism and variance of the interaction, a system based on biometrics able to work in real-time could lead to several applications in the field of socially assistive robotics, like the numerous existing application fields as autistic children, considering their well-known interest on computers and electronic devices [31], as well as people in rehabilitation in cases of dementia [39] or post-stroke [38], and generally for elderly care [4]. Moreover, a robot that bases its behavior on the recognized soft biometrics could be used to autonomously start a specific task.

Soft biometrics have been employed for filtering a large biometric database [43], as a way to improve the speed or the search efficiency of the biometric system, but they have been very rarely taken into account for HRI and artificial intelligence application. For example, in [30] user's non-verbal communication is taken into account in the design of a social robot, but a endless range of alternative biometrics could be considered for the design.

As a preliminary step, especially in order to create a fully automatic face analysis system, facial images of men and women must be extracted. The well-known Viola-Jones [41] algorithm introduces a robust cascade detector (based on AdaBoost [15] and Haar features) for the face detection in image, and is actually considered as a state-of-art approach.

Gender recognition can be viewed as a two-class classification problem, and methods can be roughly divided in feature-based and appearance-based. Mäkinen and Raisamo [29] and Sakarkaya et al. [34] introduced two wide interesting surveys that exhaustively cover the topic.The very first results were simultaneously shown in [13] and [17], in 1990. A following study, that investigated the use of geometrical features in order to achieve gender recognition, was performed by Brunelli and Poggio [7](1995), while Abdi at al. [18], in the same year, applied pixel based methods and used a radial-basis function (RBF) network. Lyons at al. used Gabor wavelets with PCA and Linear discriminant analysis (LDA) [28]. In 2002, Sun at al. showed the importance of features selection for generic algorithms [36] first and, successively, tested the efficiency of Local Binary Pattern (LBP) for gender classification [35]. Seetci at al. applied Active Apparence Models (AAM) to this scope [33], with the support of an SVM classifier. Recently, Ihsan et al. showed the performance of a Spatial Weber Local Descriptor (SWLD) [40].

Aging estimation is one of the most investigated and not trivial biometric issue. Indeed people with the same calendar age could exhibit highly different biological age because of an harder or relaxed lifestyle. AAM is a widely used techniques for age group classification used for instance by Liu at al. [24]. Doung at al. combine Active Appearance Models (AAMs) technique and LBP local facial features in combination, while Ylioinas at al. exploit the variant of LBP features to encode facial micro-patterns [44,45].

The problem of gender estimation, together with all the other information extractable from facial images, as a concept to enhance a HRI applications has been taken into account already in [42], but gender has been considered only for the design of humanoid faces, and not as a possibility of improving social interaction thanks to the possibility to perform a recognition task on the user's face. Recently, in [23], performances comparison of gender and age group recognition to carry out robot's application service for HRI has been proposed, but with the usage of audio information only. The work of [27] addresses the same problem, but using a RGB-D device and basing its processing on the body shape.

Although several works on the topic of gender recognition have been proposed over the years, in both academia and industry, it seems that very few applications of it in the field of human-robot interaction have been taken into account. Moreover, the only work of this kind in the state of the art does not explore 2D visual information. About age recognition, the work of [16] is a first attempt to process images coming from a camera installed on a robot. Authors use genetic algorithm and self-organizing maps in order to achieve the estimation, without dealing with all related problems of unconstrained environments. In the work in [26], the age estimation is given by using radial basis function and support vector machines, but very few faces are used for training and results can not be generalized.

To overcome to these limitations, this work presents a real-time system that processes data coming from a camera on board the robot to estimate gender and age group of the subjects in the scene. This way, the robot gets awareness of the interacting person and it can rapidly adapt its behavior to the subject category in order to reach the predefined assistive goal. The system can manage multiple persons at the same time, recognizing gender and age group of each person, customizing its response in each case. All privacy principles in the field of biometrics and ambient intelligence have been taken into account [6]. The manuscript is organized as follows: in section 2, our system is presented. After introducing the overall scheme, we will focus on the two used estimation algorithms. Section 3 shows experimental results. Finally, obtained results and future developments are discussed in section 4.

2 NAO biometrics based behavior system

2.1 Overview of the system

In Fig. 1 a scheme of the proposed system is shown. It is composed by two main units: the first unit is the Aldebaran NAO humanoid robot, while the second one is a Remote Computational Unit (RCU) aimed to perform all the computational tasks. RCU and NAO are connected by a local network, as shown in Fig. 1. This architecture allows to work in real-time (avoiding an overload on the low computational power of the robot's ATOM Z530 CPU), that is a fundamental requirement in the case of assistive applications.

Video frames, coming from the camera mounted on the top of the head of the robot, are taken by means of the API (Application Programming Interface) provided with the NAO Software Development Kit. Captured video frames are sent to the Biometric Engine (BE) subsystem in order to detect the presence of a human being and predict his/her gender and age group. Biometric predictions are then sent to the Behavior Decision Module (BDM) that sends a message to the robot in order to activate gender/age specific behaviors.

Communication between NAO and the RCU has been achieved using the NAOqi framework, that allows homogeneous communication between different modules (motion, audio, video), homogeneous programming and homogeneous information sharing. After connecting to the robot using an IP address and a port, it is then possible to call all the NAO's API methods as with a local method. For further information, refer to the official documentation [1].

2.2 Biometric Engine

The system core is the *biometric engine*. It uses the raw video frames as input to detect the presence of a human being and predict, through the specific *gender prediction module* and *age prediction module*, his/her gender and age group respectively. Both modules follow the scheme illustrated in Fig.2, where the first step is aimed at detecting

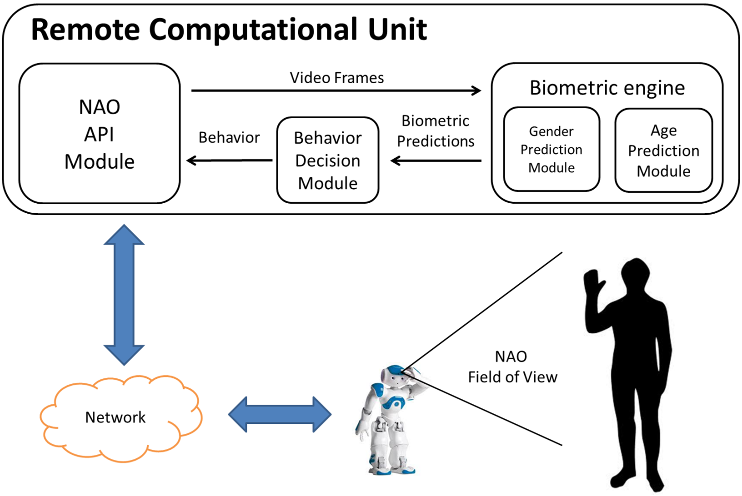


Fig. 1: A scheme of the proposed NAO biometric based behavior system.

the presence of a human face in the scene and, eventually, at cropping and normalizing the detected region. The *face detection and normalization* process is performed by means of the procedure proposed by Castrillon et al. in [9] that allows the system to detect and track also multiple faces making it, in this way, able to properly match the robot behavior to each specific person. Once the normalized face patch is available, the *features extraction* phase is performed. Taking under consideration a preliminary comparative study of various feature descriptors for gender and age group estimation [8], we choose to work with Histogram of Oriented Gradients (HOG) for gender estimation and Spatial Weber Local Descriptor (SWLD) for age problem. To achieve both biometric classifications, the `features data vector` is projected into a lower dimensional subspace through the Linear Discriminant Analysis (LDA) [3] approach. Successively, the `reduced features data vector` is given as input to the *SVM prediction* block, that predicts the biometric characteristic of interest. Both LDA and SVM approaches are trained using a dataset of thousands of faces that is freely available on line. The predicted classes are stored, frame by frame, in a circular *predicted class buffer* and finally the temporal consistency of class predictions is used to filter isolated errors: this is done by a *majority filter* that determine the gender and age group class (`predicted class`), for each person, as the class getting the greatest occurrences in the relative buffer.

Face detection and normalization The detection and normalization of the face in the scene are illustrated in Fig. 3. The normalization is a fundamental preprocessing step since the subsequent algorithm works better if they evaluate input faces with predefined size and pose. At first, the well known Viola-Jones face detector [41] is applied on the frame under investigation. When a frontal face is detected, the skin color model is used

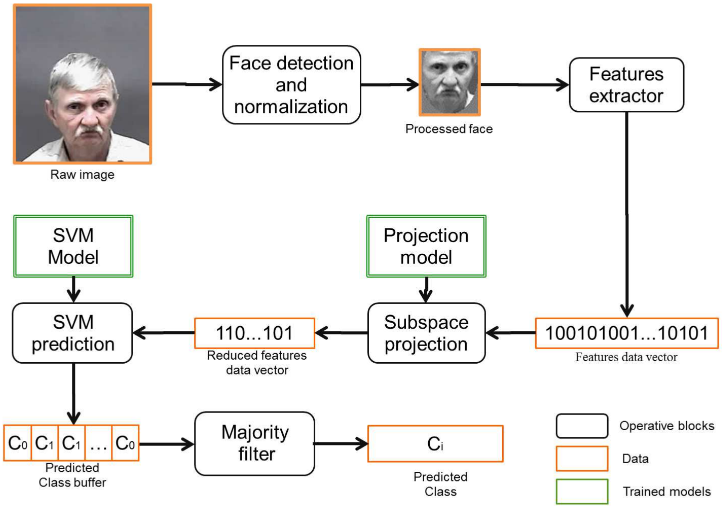


Fig. 2: The block diagram of the class prediction algorithm: the raw frames are processed in order to obtain a reliable class-prediction of the people in the scene.

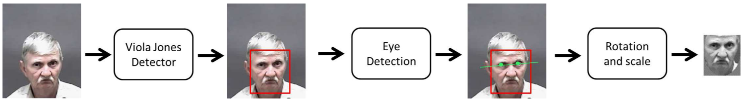


Fig. 3: The face detection and normalization step: the face is cropped and aligned in order to guarantee a standard pose to the *features extraction* step.

to detect the face blob boundaries and then the system heuristically removes elements that are not part of the face, e.g. neck, and fits an ellipse to the blob, rotating it to a vertical position. The Viola-Jones based eye detector is then applied. Eye positions, if detected, provide a measure to normalize the frontal face candidate to a standard size of 65×59 pixels.

Features extraction Extracted faces as described in the previous section are the input of the following steps. In particular, at first, the face appearance has to be encoded by using some descriptors able to emphasizes the most discriminative features for the classification problem to be faced. The choice of the descriptor is strictly related to the specific biometric information of interest as each descriptor catches a feature set that can exhibit different capabilities to discriminate different biometric aspects.

Taking under consideration a preliminary comparative study of various feature descriptors for gender and age group estimation [8], we choose to work with Histogram of Oriented Gradients (HOG) for gender estimation and Spatial Weber Local Descriptor (SWLD) for age problem.

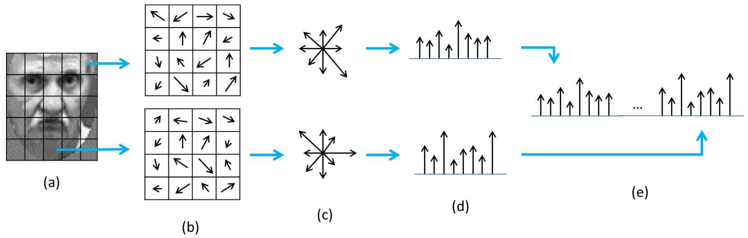


Fig. 4: HOG features extraction: the image is spatially divided in cells and the pixel orientation of each pixel in a cell is computed. Successively orientations histograms are computed and concatenated depending on the cell-space image division

HOG is a well known feature descriptor based on the accumulation of gradient directions over the pixel of a small spatial region referred as a “cell”, and in the consequent construction of a 1D histogram. Even thought HOG has many precursors, it has been used in its mature form in Scale Invariant Features Transformation [25] and widely analyzed in human detection by Dalal and Triggs [14]. This method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. Let L be the image to analyze. The image is divided in cells (Fig. 4 (a)) of size $N \times N$ pixels and the orientation θ of each pixel $x = (x_x, x_y)$ is computed (Fig. 4 (b)) by means of the following rule:

$$\theta(x) = \tan^{-1} \frac{L(x_x, x_y + 1) - L(x_x, x_y - 1)}{L(x_x + 1, x_y) - L(x_x - 1, x_y)} \tag{1}$$

The orientations are accumulated in an histogram of a predetermined number of bins (Fig. 4 (c-d)). Finally histograms of each cell are concatenated in a single spatial HOG histogram (Fig. 4 (e)). In order to achieve a better invariance to disturbs, it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measure of local histogram energy over larger spatial regions, named blocks, and using the results to normalize all of the cells in the block. The normalized descriptor blocks will represent the HOG descriptors.

Weber Local Descriptor (WLD) [11] is a robust and powerful descriptor inspired to Weber’s law. It is based on the fact that the human perception of a pattern depends not only on the amount of change in intensity of the stimulus but also on the original stimulus intensity. The proposed descriptor consist of two components: differential excitation (DE) and gradient orientation(OR). Differential excitation allow to detect the local salient pattern by means of a ratio between the relative intensity difference of the current pixel against its neighbor and the intensity of the current pixel. Moreover the OR of the single pixel is considered. When DE and OR are computed for each pixel in the image we construct a 2D histogram of T columns and $M \times S$ rows where the T is the number of orientation and $M \times S$ the number of bins for the DE quantization with the meaning of [11]. Fig. 5 showes how 2D histogram is mixed in such a way to obtain a 1D histogram more suitable for successive operation.

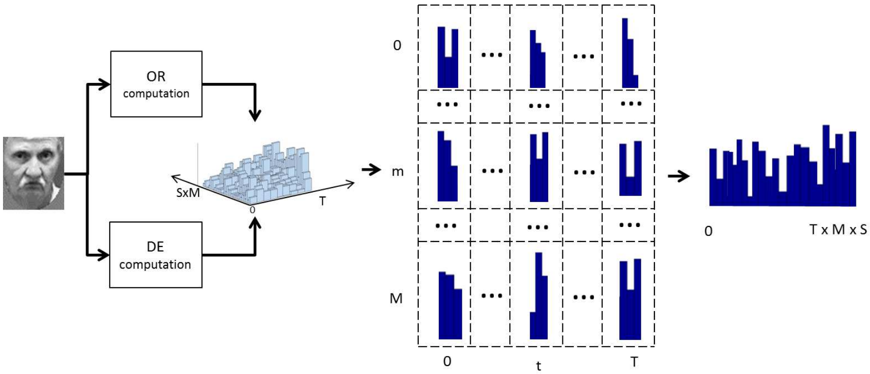


Fig. 5: WLD histogram construction process: the algorithm compute the DE and OR values for each pixel and construct the 2D histogram. The 2D histogram is split in a $M \times T$ matrix where each element is an histogram of S bins. Finally the whole 1D histogram is composed by the concatenation of the previous matrix rows.

With the aim to take into account the spatial information, we choose to use the Spatial WLD approach (SWLD) [40] that splits the image in sub-regions and to compute an histogram for each of them. Finally, histograms are concatenated in an ordered way.

Subspace projection In order to improve the computational efficiency and the accuracy of the age and gender classification, a feature reduction step is then performed. To this end, Linear Discriminant Analysis (LDA) [3] is used. In LDA, within-class and between-class scatters are used to formulate criteria for class separability. The optimizing criterion in LDA is the ratio of between-class scatter to the within-class scatter. The solution obtained by maximizing this criterion defines the axes of the transformed space. Moreover, in LDA analysis, the dimension of the projection subspace is $c - 1$, where c represents the number of class.

A more classical approach is represented by the Principal component analysis (PCA). It chooses a dimensionality reducing linear projection that maximizes the scatter of all projected samples. Simply speaking, the more informative subspace direction are selected for the subspace reduction. Anyway our tests showed that LDA, as discriminative approach, outperform the PCA. This is a straightforward result since LDA involves the labels information in the estimation of the projection model.

SVM prediction After data projection, in the proposed approach age and gender classification are performed by using Support Vector Machines (SVM). In particular two different SVMs are used: the first one is trained to discriminate among three different age groups whereas the second one is trained to discriminate between male and female human gender.

SVM is a discriminative classifier defined by a separating hyperplane. Given a set of labeled training data (supervised learning), the algorithm computes an optimal hyperplane (the trained model) which categorizes new examples in the right class. In particular the C-support vector classification (C-SVC) learning task implemented in the well-known LIBSVM [10] has been used. Given training vectors $\boldsymbol{x}_i \in \mathbb{R}^n, i = 1, \cdots, l$, in two classes, and an label vector $\boldsymbol{y} \in \mathbb{R}^l$ such that $y_i \in \{1, -1\}$, C-SVC [5, 12] solves the following primal optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w}, b, \xi} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \cdots, l, \end{aligned}$$

where $\phi(\boldsymbol{x}_i)$ maps $\boldsymbol{x}_i$ into a higher-dimensional space and $C > 0$ is the regularization parameter. Due to the possible high dimensionality of the vector variable $\boldsymbol{w}$, usually the following dual problem is solved:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}^T Q\boldsymbol{\alpha} - \boldsymbol{e}^T\boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, \cdots, l, \end{aligned}$$

where $\boldsymbol{e} = [1, \cdots, 1]^T$ is the vector of all ones, Q is an $l \times l$ positive semidenite matrix, $Q_{ij} y_i y_j K(\boldsymbol{x}_i, , \boldsymbol{x}_j)$, and $K(\boldsymbol{x}_i, , \boldsymbol{x}_j) \equiv \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$ is the kernel function. After the dual problem is solved, the next step is to compute the optimal w as:

$$\boldsymbol{w} = \sum_{i=1}^{l} y_i\alpha_i\phi(\boldsymbol{x}_i) \tag{2}$$

Finally the decision function is

$$\mathtt{sgn}(\boldsymbol{w}^T\phi(\boldsymbol{x}) + b) = \mathtt{sgn}\left(\sum_{i=1}^{l} y_i\alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \tag{3}$$

Such an approach is suitable only for the two classes gender problems. For age group classification a multi-class approach has been used. It can be treated through the "one-against-one" [22]. Let k be the number of classes, then $k(k-1)/2$ classifiers are constructed where each one trains data from two classes. The final prediction is returned by a voting system among all the classifiers. Many other methods are available for multi-class SVM classification, anyway in [20] a detailed comparison is given with the conclusion that "one-against-one" is a competitive approach.

3 Experimental results

The experimental session consists of two main steps. The first one aims at evaluating the *biometric prediction accuracy* and it has been performed off-line (without the NAO) over a large set of faces collected in a database and it allows to validate the robustness of age and gender classification algorithms. Successively, *on line tests* involving the complete framework (NAO + RCU) have been performed in order to validate the HRI environment.

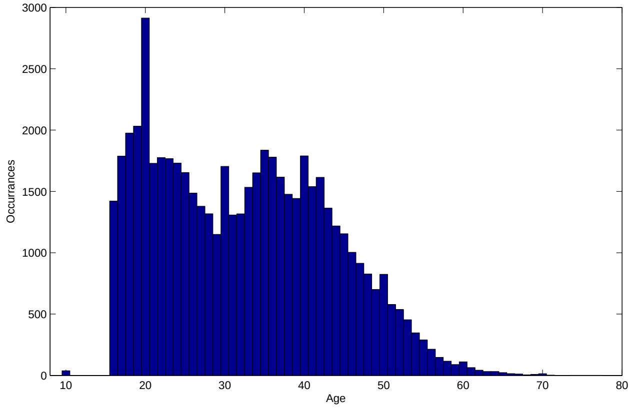


Fig. 6: Histogram for the age distribution in the Fusion dataset: the age is highly unbalancend toward the mean age range and this can affects the prediction accuracy.

3.1 Off-line Biometric prediction accuracy estimation

The evaluation of the accuracy of both *gender prediction module* and *age prediction module* has been realized with a k-fold test over the whole model estimation and prediction process. Experiments have been carried out on a fusion of two of the most representative datasets used for face classification problems: the Morph [2] and the Feret [32] datasets. Both datasets consist of face images of people of different gender, ethnicity and age and to each image is associated a descriptive CVS file containing related information. The total number of face images used in the experimental phase has been 65341 (55915 males and 9246 females). The age distribution in the considered dataset is represented in Figure 6.

The accuracy estimation of the proposed prediction algorithms has been performed following the procedure, showed in Fig.7. It consists of two steps: a model estimation and a prediction estimation. The whole dataset has been randomly split in k sub-folds. For each of the k validation steps, $k-1$ sub-fold for the training and 1 sub-fold for the prediction/validation process have been used. Face detection and normalization has been the performed on each image of the selected k-1 training sub-fold and then the `features data vectors` have been extracted. The set of `features data vectors` has been then used to train, in sequence, the feature reduction algorithm and the SVMs. Finally the one-out fold has been finally tested by using the available models.

The process is repeated over each of 5 to one-out sub-fold combination and the accuracy results is averaged.

For HOG operator, the *VLFeat library*[1] has been used using standard parameters as in [14] that lead to a features vector of 2016 elements. On the other hand the SWLD

[1] http:www.vlfeat.org

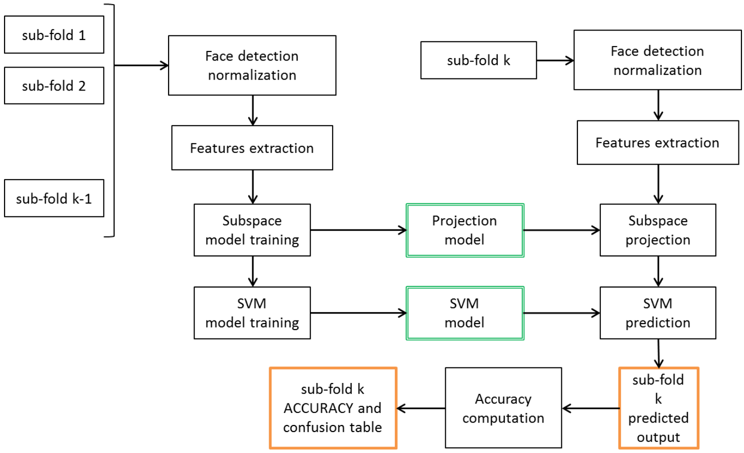


Fig. 7: Test procedure for accuracy estimation: the procedure is done k times in order to obtain the best estimation of total accuracy and confusion table.

operator has been developed following the line of [40] and used with value of T, M, S that are respectively of 8, 4 and 4 over a 4×4 grid of the image with a final features vector of 2048 elements.

The SVM classification problem has been treated by means of the publicly available LIBSVM library [10]. More precisely, we used a radial basis function (RBF) that, in the opinion of the authors of [19] as well as in our experience, seems to be the most reasonable choice. Usually a grid search for penalty parameters C and the others RBF parameters could be desirable. Anyway, our tests does not arise any significant difference in the results as the parameters change. More specifically, we set $C = 1$ and $\gamma = 1/N_f$ where N_f is the number of features.

The class separation for gender task is straightforward. On the other hand, the ages have been split in three categories: *young* (people with an age $\leq$25), *middle* (people with an age $>$25 and $\leq$50) and *old* (people with an age $>$50). This splitting was thought in order to take into account generics assistive applications, as well as general purpose customized activities during rehabilitation.

We obtained a total accuracy of 88.6% for gender classification (developed with the HOG descriptor) and 62.1% for age group classification. Moreover the results are presented in form of confusion tables (see Table 1) that allow a clear point view of the real performance of each estimator where M, F, Y, M, O, mean male, female, young, middle and old respectively the subfixs P and T stay for predicted and true. About gender recognition performances, it is quite evident that the imbalance among male and female true positive rate is a consequence of the insufficient female entries on the dataset.

Table 1: Confusion table for gender (a) and age group (b) estimation (k-fold = 5)

(a)

	M_P	F_P
M_T	98.7%	1.3%
F_T	21.5%	78.5%

(b)

	Y_P	M_P	O_P
Y_T	69.4%	30.6%	0%
M_T	13.0%	85.2%	1.8%
O_T	0.6%	67.5%	31.9%

Table 2: Behavior table: depending on the predicted gender and age group, the robot reacts with a specific and predefined behavior.

	M	F
Y_T	High five+*Hi bro*	High five+*Hi sister*
M_T	Hello	Bow down
O_T	Handshake	Hand clap

3.2 On line tests of the whole framework

The whole architecture (presented in section 2) has been tested also in a real scenario where people directly interacted with the robot. No constraints about appearance nor background were given to the participants. Persons entered the scene in the field of view of the NAO robot and when their faces were detected, the algorithms running on the remote processing units predicted their gender and age group and consequently the robot acted in the proper way. For testing purposes, i.e. in order to show the actual possibility to adapt the robot's behavior to the the user biometrics, a set of behaviors was defined and associated to the predicted biometrics as summed up in Table 2.

For example, first row shows the behavior to be taken in the presence of a male or a woman classified as young: in both cases, the robot highs five, but in case of man, it says the sentence ”Hello bro!”, while in case of a girl, the sentence is ”Hello sister!”. The behaviors related to the case of predicted presence of middle aged people are instead reported in the second row. In this case, if a woman is in the scene, NAO bows down, while in the presence of a man, the robot greets with his right hand. Fig. 8 illustrates the NAO point of view and the recognition step and the subsequent action taken depending on the gender prediction. Finally, the third row shows the behavior to be taken in the case of older people: in the case of a man, the robot proposes an handshake to the user, otherwise it starts the hand clap game. Even a sentence to be pronounced by the robot has been customized depending on the sex and the age. In the presence of a mixed group, when a detection occur, the robot can say the exact number of men and women in the scene, as well as the age group of each participant. Since given a person each prediction is independent from the possible presence of other faces in the same image, it was possible to estimate the error of the system evaluating the interaction with the robot of one person at time. With our real scenario, we tested the algorithm on 20 persons. More precisely, 10 men and 10 women participated to the experiment covering all the age groups accounted in the experiment (from 20 to 65 years old). The test reports 3 errors for the gender, while 4 misclassifications occurred in the case of age.

About computational remarks, the system was tested on a local network in order to avoid latency errors in the evaluation of the frame rate. The RCU was a CPU i7@3.20GHz with a RAM of 16 GB DDR3. Images were processed as a resolution of 640×480. In these conditions, our system was able to work at a frame rate of 13 fps. This is a very encouraging result since it allowed to use the predicted gender buffer in order to strengthen the prediction.

4 Conclusions

With this work, the idea of using biometrics as a way to improve human-robot interaction has been investigated. In particular, gender and age estimation algorithms have been used to provide awareness of the user biometrics to a humanoid robot (Aldebaran NAO), in order to properly react with a specific gender/age behavior. The system can also manage multiple persons at the same time, recognizing the age and gender of each participant. The system is particularly suited for HRI applications requiring a natural level of interaction like training, rehabilitation, home-care and so on. Future works will further investigate the use of more distinctive features in order to improve the accuracy in gender and age group classification, as well as for the face detection algorithm in order to overcome the state of the art drawbacks concerning the handling of non-frontal face pose and of the strong illumination variations. About the unbalanced performance on the gender recognition task, a balanced dataset will be employed. Finally, a larger set of robot's behaviors is going to be developed in conjunction with a statistical model able to take decision about the mapping between the outcomes of the algorithmic procedures and the robot actions.

References

1. (May 2014), https://community.aldebaran-robotics.com/doc/1-14/index.html
2. (May 2014), http://www.faceaginggroup.com/morph/
3. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on 19(7), 711–720 (Jul 1997)
4. Bemelmans, R., Gelderblom, G.J., Jonker, P., De Witte, L.: Socially assistive robots in elderly care: A systematic review into effects and effectiveness. Journal of the American Medical Directors Association 13(2), 114–120 (2012)
5. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. pp. 144–152. COLT '92, ACM, New York, NY, USA (1992), http://doi.acm.org/10.1145/130385.130401

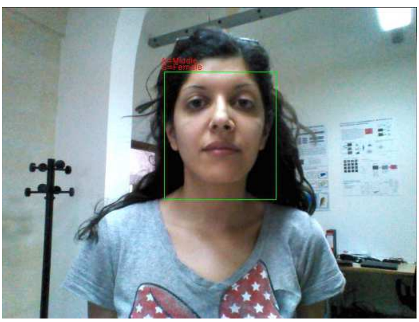
6. Brey, P.: Freedom and privacy in ambient intelligence. Ethics and Information Technology 7(3), 157–166 (2005)
7. Brunelli, R., Poggio, T.: Hyberbf networks for gender classification (1995)
8. Carcagni, P., Del Coco, M., Mazzeo, P., Testa, A., Distante, C.: Features descriptors for demographic estimation: a comparative study. In: Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage (VAAM). In Conjunction with the 22nd International Conference on Pattern Recognition ICPR 2014 (Aug 2014), accepted for pubblication
9. Castrillón, M., Déniz, O., Guerra, C., Hernández, M.: Encara2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation 18(2), 130–140 (2007)
10. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
11. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: Wld: A robust local image descriptor. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(9), 1705–1720 (Sept 2010)
12. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995), `http://dx.doi.org/10.1023/A%3A1022627411411`
13. Cottrell, G.W., Metcalfe, J.: Empath: Face, emotion, and gender recognition using holons. In: Advances in Neural Information Processing Systems. pp. 564–571 (1990)
14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893 vol. 1 (June 2005)
15. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55(1), 119–139 (Aug 1997)
16. Fukai, H., Nishie, Y., Abiko, K., Mitsukura, Y., Fukumi, M., Tanaka, M.: An age estimation system on the aibo. In: Control, Automation and Systems, 2008. ICCAS 2008. International Conference on. pp. 2551–2554 (Oct 2008)
17. Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: Sexnet: A neural network identifies sex from human faces. In: NIPS. pp. 572–579 (1990)
18. H., A., D., V., B., E., AJ., O.: More about the difference between men and women: evidence from linear neural networks and the principal-component approach. Neural Comput. 7(6), 1160 – 1164 (1995)
19. Hsu, C.W., Chang, C.C., Lin, C.J., et al.: A practical guide to support vector classification (2003)
20. Hsu, C.W., Lin, C.J.: A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions on 13(2), 415–425 (Mar 2002)
21. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: Biometric Authentication, pp. 731–738. Springer (2004)
22. Knerr, S., Personnaz, L., Dreyfus, G.: Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Souli, F., Hrault, J. (eds.) Neurocomputing, NATO ASI Series, vol. 68, pp. 41–50. Springer Berlin Heidelberg (1990), `http://dx.doi.org/10.1007/978-3-642-76153-9_5`
23. Lee, M.W., Kwak, K.C.: Performance comparison of gender and age group recognition for human-robot interaction. International Journal of Advanced Computer Science & Applications 3(12) (2012)
24. Liu, L., Liu, J., Cheng, J.: Age-group classification of facial images. In: Machine Learning and Applications (ICMLA), 2012 11th International Conference on. vol. 1, pp. 693–696 (Dec 2012)
25. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (Nov 2004)

26. Luo, R., Chang, L.W., Chou, S.C.: Human age classification using appearance images for human-robot interaction. In: Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE. pp. 2426–2431 (Nov 2013)
27. Luo, R.C., Wu, X.: Real-time gender recognition based on 3d human body shape for human-robot interaction. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction. pp. 236–237. ACM (2014)
28. Lyons, M.J., Budynek, J., Plante, A., Akamatsu, S.: Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis. In: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. pp. 202–207 (2000)
29. Mäkinen, E., Raisamo, R.: An experimental comparison of gender classification methods. Pattern Recognition Letters 29(10), 1544 – 1556 (2008), `http://www.sciencedirect.com/science/article/pii/S0167865508001116`
30. McColl, D., Zhang, Z., Nejat, G.: Human body pose interpretation and classification for social human-robot interaction. International Journal of Social Robotics 3(3), 313–332 (2011)
31. Moore, D.: Computers and people with autism. Asperger Syndrome pp. 20–21 (1998)
32. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The feret evaluation methodology for face-recognition algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(10), 1090–1104 (Oct 2000)
33. Saatci, Y., Town, C.: Cascaded classification of gender and facial expression using active appearance models. In: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on. pp. 393–398 (April 2006)
34. Sakarkaya, M., Yanbol, F., Kurt, Z.: Comparison of several classification algorithms for gender recognition from face images. In: Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on. pp. 97–101 (June 2012)
35. Sun, N., Zheng, W., Sun, C., Zou, C., Zhao, L.: Gender classification based on boosting local binary pattern. In: Wang, J., Yi, Z., Zurada, J., Lu, B.L., Yin, H. (eds.) Advances in Neural Networks - ISNN 2006, Lecture Notes in Computer Science, vol. 3972, pp. 194–201. Springer Berlin Heidelberg (2006)
36. Sun, Z., Bebis, G., Yuan, X., Louis, S.J.: Genetic feature subset selection for gender classification: A comparison study. In: In: IEEE Workshop on Applications of Computer Vision. pp. 165–170 (2002)
37. Tapus, A., Maja, M., et al.: Towards socially assistive robotics. International journal of the Robotics Society of Japan (JRSJ) 24(5), 576–578 (2006)
38. Tapus, A., Ţăpuş, C., Matarić, M.J.: User?robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intelligent Service Robotics 1(2), 169–183 (2008)
39. Tapus, A., Tapus, C., Mataric, M.J.: The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In: Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on. pp. 924–929. IEEE (2009)
40. Ullah, I., Hussain, M., Muhammad, G., Aboalsamh, H., Bebis, G., Mirza, A.: Gender recognition from face images with local wld descriptor. In: Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on. pp. 417–420 (April 2012)
41. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. vol. 1, pp. I–511. IEEE (2001)
42. Walker, J.H., Sproull, L., Subramani, R.: Using a human face in an interface. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 85–91. ACM (1994)
43. Wayman, J.L.: Large-scale civilian biometrics system - issues and feasibility. In: Proceedings of the CardTech/SecureTech Government, Washington DC (1997)

44. Ylioinas, J., Hadid, A., Pietikainen, M.: Age classification in unconstrained conditions using lbp variants. In: Pattern Recognition (ICPR), 2012 21st International Conference on. pp. 1257–1260 (Nov 2012)
45. Ylioinas, J., Hadid, A., Hong, X., Pietikinen, M.: Age estimation using local binary pattern kernel density estimate. In: Petrosino, A. (ed.) Image Analysis and Processing ?? ICIAP 2013, Lecture Notes in Computer Science, vol. 8156, pp. 141–150. Springer Berlin Heidelberg (2013)

(a) NAO gender recognition step (male/middle).

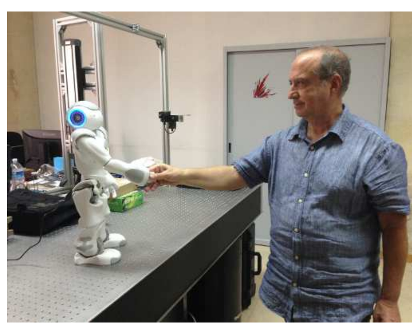
(b) Behavior after recognition (male/middle).



(c) NAO gender recognition step (female/middle).

(d) Behavior after recognition (female/middle).

(e) NAO gender recognition step (male/mature).



(f) Behavior after recognition (male/mature).

Fig. 8: A test of the interaction between the NAO and humans being. The NAO recognizes gender and age group of the interacting subject (a,c,e) and reacts with a customized behavior (it prostrates for middle aged woman, it greets with its right hand in the presence of middle aged male and shakes hand with a mature man).