DOI: 10.1109/BIOMS.2012.6345775

Weight Estimation from Frame Sequences Using Computational Intelligence Techniques

Ruggero Donida Labati, Angelo Genovese, Vincenzo Piuri, Fabio Scotti
Department of Information Technology
Università degli Studi di Milano
Milano, 20122, Italy.
{ruggero.donida, angelo.genovese, vincenzo.piuri, fabio.scotti}@unimi.it

***Abstract*—Soft biometric techniques can perform a fast and unobtrusive identification within a limited number of users, be used as a preliminary screening filter, or combined in order to increase the recognition accuracy of biometric systems. The weight is a soft biometric trait which offers a good compromise between distinctiveness and permanence, and is frequently used in forensic applications. However, traditional weight measurement techniques are time-consuming and have a low user acceptability. In this paper, we propose a method for a contactless, low-cost, unobtrusive, and unconstrained weight estimation from frame sequences representing a walking person. The method uses image processing techniques to extract a set of features from a pair of frame sequences captured by two cameras. Then, the features are processed using a computational intelligence approach, in order to learn the relations between the extracted characteristics and the weight of the person. We tested the proposed method using frame sequences describing eight different walking directions, and captured in uncontrolled light conditions. The obtained results show that the proposed method is feasible and can achieve a view-independent weight estimation, also without the need of computing a complex model of the body parts.**

***Index Terms*—soft biometrics, weight, neural networks.**

I. Introduction

Soft biometric recognition techniques, while featuring a lack of distinctiveness, can use samples captured in an unobtrusive and unconstrained way, in uncooperative conditions [1–3], or with surveillance cameras placed at long distances [3]. Such recognition systems can be employed where it is difficult to adopt systems based on hard biometric traits (e.g. surveillance applications), the pool of users is small enough, or a high accuracy is not required. Examples of soft biometric traits are the height, gender, and eye color.

In this context, the weight offers a good compromise between distinctiveness and permanence [4]. Moreover, in forensic analyses, the weight is one of the few characteristics that can be inferred from the evaluation of a scene. Traditional techniques for the weight measurement, however, are based on the contact of the body with a sensor and are difficult to be applied in uncooperative contexts. The contactless weight estimation based on surveillance frame sequences can reduce the time needed for the measurement process, increase the user acceptability, and be a useful tool in investigative activities. Other possible fields of application are the surveillance of critical areas, public buildings, schools (e.g., against pedophiles), and public areas (e.g., for safety monitoring).

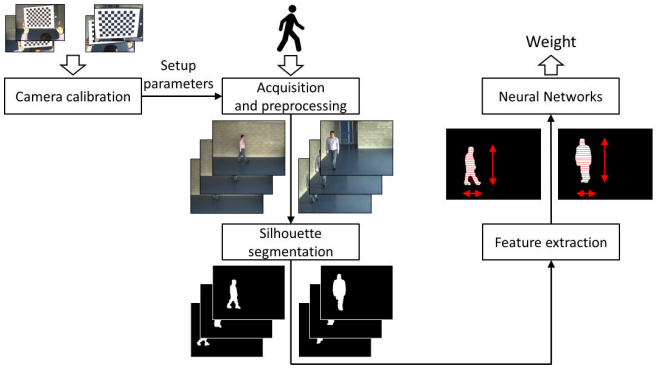

Fig. 1. Schema of the proposed method.

In this paper, we propose a method for a contactless, low-cost, and unobtrusive weight estimation based on frame sequences captured in unconstrained environments. The method achieves a weight estimation independent from the point of view, position of the person, and illumination. Moreover, the proposed method does not require the computation of complex models of the body parts. The schema of the proposed method is shown in Fig.1. First, the method performs the analysis of pairs of frame sequences captured by two cameras, and extracts features related to the dimensional characteristics of the silhouette. A computational intelligence approach is then used to process the features and estimate the corresponding weight, by evaluating the relations between the visual characteristics and the weight of the person.

The paper is structured as follows. Section II describes the related works on weight estimation techniques and soft biometric systems based on surveillance cameras. In Section III, the proposed method is detailed. Section IV presents the experimental results, and Section V summarizes the work.

II. Previous work

In the literature, recognition techniques designed for surveillance applications are based on different biometric traits. There are methods based on hard-biometric traits, for example the face characteristics [5,6], and methods based on soft biometric traits [3]. The main advantages of using soft biometric techniques consist in the possibility to perform the recognition in less-constrained scenarios, and in the opportunity to design fast techniques for the continuous authentication and periodic re-authentication. Soft biometric traits used in surveillance ap-

plications can be behavioral or physiological. In the literature, the most studied behavioral soft biometric trait is the gait [7]. Gait recognition systems can obtain a satisfactory accuracy and can work at great distances [8]. Physiological soft biometric traits can be used in different applicative contexts. The method described in [9] computes categorical information about an individual (e.g. gender and race) in order to filter large surveillance databases by limiting the number of entries to be searched for each biometric query. The approach proposed in [10] uses characteristics extracted from the face and the clothes in order to perform continuous authentications. A method based on color and height characteristics for the detection of the individuals throughout a sparse multi-camera network is presented in [11]. The technique described in [3] uses three part (head, torso, legs) height and color soft biometric models in order to perform the recognition of the individuals. Techniques for the extraction of characteristics related to the gait, height, size, and gender are presented in [12].

A critical step for many biometric applications based on surveillance cameras is the silhouette segmentation. In the literature, there are many studies on silhouette segmentation techniques, which can be divided in methods based on the direct detection and methods based on the background subtraction. Examples of methods based on the direct detection are presented in [13,14]. With respect to the methods based on the direct detection, most of the methods based on the background subtraction can obtain better performances in terms of computational time. For this reason, these methods are more used in surveillance applications than direct detection methods. Some well-known background subtraction techniques in the literature are described in [15–17]. The presence of shadows can drastically reduce the segmentation accuracy of these methods, especially in surveillance applications with uncontrolled light conditions. In the literature, there are many algorithms for the shadow removal [18], which consider different features (e.g. the gradient of the images, the correlation between frames, and models of the light diffusion).

An interesting soft biometric characteristic is the weight. This trait, in fact, offers a good compromise between distinctiveness and permanence and can be used in forensic applications. In the literature, there are few studies on the weight estimation from the visual aspect of the individuals. The correlation between the human metrology and the weight is studied in [19]. The method described in [4] estimates the weight from single images by using measurements performed by skilled operators. These techniques require the support of a human expert and do not consider walking persons.

III. The proposed approach

The proposed approach is designed to estimate the weight of a walking person by using image processing techniques and a computational intelligence approach. The method performs a contactless, low-cost, unobtrusive, and unconstrained weight estimation, without defining and computing complex relationships between the size of various body parts and the body weight. The proposed weight estimation technique is invariant to the point of view, illumination conditions, position, and speed of the walking person. Satisfactory results were obtained using a similar approach for the estimation of the volume of small objects [20].

The weight is estimated from frame sequences captured by two cameras placed in order to obtain a frontal view and a side view of the walking person. The method extracts a set of dimensional features from the silhouette of the individual. Then, a first approximation of the body volume is computed, and a technique based on neural networks is used to perform the weight estimation.

The approach can be divided in the following steps:

1) camera calibration;
2) acquisition and preprocessing;
3) silhouette segmentation;
4) feature extraction;
5) weight estimation using neural networks.

A. Camera calibration

The cameras are individually calibrated off-line before the acquisitions by using a set of images representing a chessboard moved in different positions. The intrinsic parameters are then computed by using the techniques proposed in [21,22]. The considered extrinsic parameters are the positions and angles of the two cameras, which are measured by a human operator.

B. Acquisition and preprocessing

Two frame sequences are captured synchronously from two color CCD cameras with wide-angle lenses. Every frame is then rectified in order to compensate for the lens distortion, using the algorithms described in [21,22]. In order to simplify the segmentation step, a reference background image is captured for every frame sequence.

C. Silhouette segmentation

The silhouette segmentation is a critical step of the proposed method. The considered frame sequences, in fact, are captured in uncontrolled light conditions and can present strong shadows. The proposed technique is based on a background subtraction approach and estimates the foreground for each single frame i by considering separately captured background images. Shadow removal techniques are used in order to overcome problems related to the uncontrolled light conditions. The technique can be divided in the following steps: estimation of the moving regions; estimation of the strong shadows; region of interest (ROI) computation; boundary estimation; computation of the correlation image; boundary filling; silhouette refinement.

The estimation of the moving regions permits to obtain a first approximation of the ROI. In order to detect the moving regions with colors similar to the background, the maximum distance between the RGB channels of the frame $I_F(i)$ and the background $I_B(j)$ is computed:

$$D(x,y,i) = \max\left(\left|c\left(I_F(x,y,i)\right) - c\left(I_B(x,y,i)\right)\right|\right), \quad (1)$$

with $c = R, G, B$. A binary image representing the moving regions $I_M(i)$ is then obtained by applying an empirically

estimated threshold t_M corresponding to the k^{th} percentile of the histogram of $D(i)$:

$$I_M(x,y,i) = \begin{cases} 1 & \text{if } D(x,y,i) < t_M \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

Strong shadows are then searched in order to prevent the presence of artifacts in the final silhouette image. Similarly to the algorithm described in [23], the channel H of the HSV colorspace is considered. First, the frame $I_F(i)$ and the background $I_B(i)$ are converted in the HSV colorspace. The binary image $I_S(i)$ describing the strong shadows is computed as:

$$I_S(x,y,i) = \begin{cases} 1 & \text{if } (t_{H1} < H_F(x,y,i) < t_{H2}) \\ & \quad \wedge (t_{D1} < D_H(x,y,i) < t_{D2}) \\ 0 & \text{otherwise} \end{cases} , \quad (3)$$

where $H_F(i)$ is the channel H of the frame $I_F(i)$ converted in the HSV colorspace, $D_H(i)$ is the angular distance between the channel H of the considered frame and the background, t_{H1}, t_{H1}, t_{H1}, and t_{H1} are empirically estimated thresholds.

The ROI is computed considering the previously estimated binary masks and the module of the gradient $G_M(i)$ of the matrix $D(i)$. First, a binary image $I_R(i)$ is computed as:

$$I_R(x,y,i) = I_M(x,y,i) \times (\neg I_S(x,y,i)) . \quad (4)$$

An image of the gradient $I_G(i)$ of the candidate ROI regions is computed as:

$$I_G(x,y,i) = I_R(x,y,i) \times G_M(x,y,i) . \quad (5)$$

The ROI is defined as the 8-connected region of $I_R(i)$ with the maximum value obtained by summing the correspondent region of $G_M(i)$. Finally, a morphological closing operation is applied to the ROI.

The boundary is then estimated in order to obtain a segmented area that is not affected by weak shadows. Weak shadows, in fact, do not present strong edges [24]. The boundary is computed by using the information related to the gradient of the considered frame and background. The binary images of the edges of the frame $E_F(i)$ and background $E_B(i)$ are obtained by applying an empirically estimated threshold t_N to the gradient module of $D(i)$ and $I_B(i)$ respectively. The boundary image $B_B(i)$ is defined as:

$$B_B(x,y,i) = E_F(x,y,i) - (\neg E_B(x,y,i)) . \quad (6)$$

The obtained result is then refined by applying a morphological opening operation followed by a morphological closing operation.

Considering that the boundary image $B_B(i)$ does not describe the complete body shape, additional information is estimated in order to properly segment the human silhouette. The image regions appertaining to the silhouette are then estimated by computing the correlation between the considered frame and background. A matrix $C(i)$ is computed as the correlation between local $m \times m$ regions of the frame $I_F(i)$ and the background $I_B(j)$ converted in gray-scale. A binary image $C_T(i)$ is obtained as

$$C_T(i,j) = \begin{cases} 1 & \text{if } C(x,y,i) < t_T \\ 0 & \text{otherwise} \end{cases} , \quad (7)$$

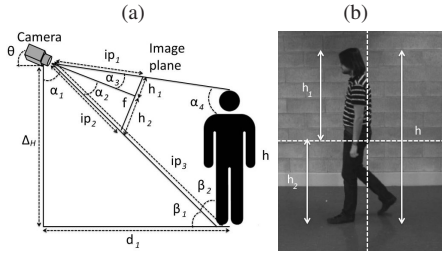

Fig. 2. Visual representation of the height computation method: (a) geometric parameters; (b) body measures.

where t_T is an empirically estimated value.

A binary image of the candidate body regions $C_B(i)$ is then obtained as:

$$C_B(x,y,i) = C_T(x,y,i) \times (\neg B_B(x,y,i)) . \quad (8)$$

In order to reduce the probability of considering regions affected by shadows as appertaining to the human silhouette, a trapezoidal area describing the body limits is estimated and only the local regions of the image $C_B(i)$ appertaining to this area are used to estimate the final silhouette. The 8-connected regions of $C_B(i)$ are considered as appertaining to the body if the $p\%$ of their area is included in the trapezoidal region describing the body limits. The limits of the trapezoidal area along the y direction correspond to the maximum and minimum coordinates of the pixels equal to 1 in the boundary image $B_B(i)$. The limits along the x direction correspond to the maximum and minimum coordinates of the pixels equal to 1 in the areas of boundary image $B_B(i)$ included between the 0% and 20% of the boundary height (corresponding to the foots) and between the 80% and 100% of the boundary height (corresponding to the head). A first approximation of the silhouette $S_M(i)$ is obtained by summing the obtained areas and the boundary $B_B(i)$. The silhouette $S(i)$ is then computed as the biggest 8-connected region of $S_M(i)$.

In order to refine the obtained result, a morphological filling operation followed by a morphological closing operation are applied.

D. Feature extraction

A set of features are computed for each pair of frame sequences by using characteristics extracted from every frame:

- the height of the person;
- an approximation of the body volume;
- a set of values describing the areas of the ellipses that approximate the body shape at different heights;
- the walking direction (estimated by a human operator).

1) Height estimation: the body height is estimated for each frame i by using a technique similar to the method presented in [25]. The proposed algorithm is based on the pinhole camera model and trigonometric equations.

The used extrinsic parameters are the height Δ_H and the tilt angle θ of the cameras (Fig. 2). The used intrinsic parameter is the focal length f, obtained from the calibration step.

The distances of the head and feet from the center of the silhouette image $S(i)$ are first extracted, obtaining $h_1(i)$ and $h_2(i)$. The values of $ip_1(i), ip_2(i)$ are then obtained as:

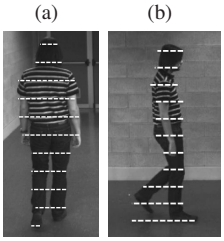

Fig. 3. Subdivision of the segmented person and computation of the width from each subdivision: (a) frame captured by Camera A; (b) frame captured by Camera B.

$$ip_1(i) = \sqrt{f^2 + h1(i)^2} \;\; ; \;\; ip_2(i) = \sqrt{f^2 + h2(i)^2} \;\; . \quad (9)$$

The angles $\alpha_2(i), \alpha_3(i)$ are computed as:

$$\alpha_2(i) = \sin^{-1}\frac{h_2(i)}{ip_2(i)} \;\; ; \;\; \alpha_3(i) = \sin^{-1}\frac{h_1(i)}{ip_1(i)} \;\; . \quad (10)$$

The angle $\alpha_1(i)$ is computed as follows:

$$\alpha_1(i) = 90^\circ - \theta - \alpha_2(i) \;\; . \quad (11)$$

The values of $d_1(i)$ and $ip_3(i)$ are then computed as:

$$d_1(i) = \Delta_H \tan\alpha_1(i) \;\; ; \;\; ip_3(i) = \sqrt{\Delta_H^2 + d_1(i)^2} \;\; . \quad (12)$$

The angles $\beta_1(i), \beta_2(i)$ are obtained as:

$$\beta_1(i) = 90^\circ - \alpha_1(i) \;\; ; \;\; \beta_2(i) = 90^\circ - \beta_1(i) \;\; , \quad (13)$$

and then the angle $\alpha_4(i)$ is computed as:

$$\alpha_4(i) = 180^\circ - \beta_2(i) - (\alpha_2(i) + \alpha_3(i)) \;\; . \quad (14)$$

Finally, the height h_i of the i-th frame is estimated as:

$$h(i) = ip_3(i)\frac{\sin(\alpha_2(i) + \alpha_3(i))}{\sin\alpha_4(i)} \;\; . \quad (15)$$

2) Volume approximation: a volume approximation is computed from each pair of frames captured at the instant i by the two cameras. The segmented silhouettes $S_A(i)$ and $S_B(i)$ are first divided in a certain number of intervals n_v along the y axis, as shown in Fig. 3.

For each height interval, an ellipse passing from the coordinates of the silhouette in the images is computed. The values $s_A(i,n)$ and $s_B(i,n)$, describing the lengths in pixel of the interval n in the silhouettes $S_A(i)$ and $S_B(i)$, are extracted. These values are then converted in millimeters by using the information related to the body height:

$$\begin{aligned} l_A(i,n) &= (s_A(i,n) \times h_A(i)) / h_{Ap}(i); \\ l_B(i,n) &= (s_B(i,n) \times h_B(i)) / h_{Bp}(i), \end{aligned} \quad (16)$$

where $h_A(i)$ and $h_B(i)$ are the height values estimated from the silhouettes $S_A(i)$ and $S_B(i)$ expressed in mm, $h_{Ap}(i)$ and $h_{Bp}(i)$ are the height values expressed in pixel, $l_A(i,n)$ and $l_B(i,n)$ are the lengths of the interval n, expressed in millimeters. The area of each ellipse is then computed as:

$$A(i,n) = l_A(i,n)^2 \, l_B(i,n)^2 \, \pi \;\; , \quad (17)$$

where $A(i,n)$ is the area of the n-th ellipse relative to a pair of frames. Finally, the volume approximation of the i-th pair of frames is computed as:

$$V(i) = \sum_{n=1}^{n_v} a(i,n) \left(\frac{h(i)}{n_v}\right). \quad (18)$$

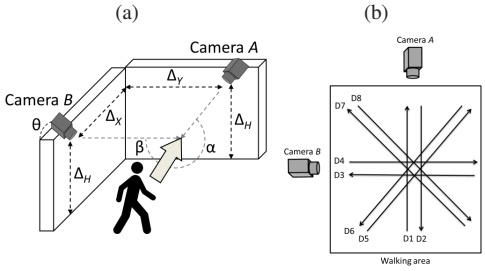

Fig. 4. Schema of the proposed acquisition setup (a) and the different walking paths (b).

E. Weight estimation using Neural Networks

Weight estimation techniques based on accurate measurements of the body parts are difficult to be performed in frame sequences of walking persons. The body, in fact, can assume different postures and some body parts can be occluded. Moreover, shadows and reflections can reduce the accuracy of the silhouette segmentation algorithms, especially in unconstrained light conditions.

In this case, only the approaches based on computational intelligence techniques are realistically feasible. In particular, the capability of the neural networks to correctly map input-output relationships starting from an example dataset can be here exploited to create a weight estimation system. The generalization capability of neural networks, moreover, can permit to obtain accurate results also using noisy input data.

Feedforward neural networks are then used to estimate the weight of the walking person by processing the vector of extracted features from a pair of frame sequences captured by two cameras.

As described in Section IV, we used feature vectors composed by 7, 17, 27, and 37 characteristics for Dataset 1, Dataset 2, Dataset 3, and Dataset 4, respectively. The obtained result consists in the estimated weight expressed in kg.

IV. Experimental Results

The proposed method has been tested on frame sequences captured in our laboratory since, at the best of our knowledge, there are not available any public datasets of surveillance frame sequences reporting the weight of the individuals. The acquisition setup used for capturing the subjects is composed by two synchronized Sony XCD-SX90CR CCD color cameras, placed at a height $\Delta_H = 2000$ mm, and positioned in order to capture both a frontal and side view of the subject (Fig. 4a). The tilt angle of the cameras with respect to the floor is $\theta = 25^\circ$, the distances between the cameras are $\Delta_X = 4360$ mm, $\Delta_Y = 5530$ mm. The cameras were individually calibrated.

The parameters used for performing the silhouette segmentation are: $k = 80$; $t_{H1} = 30$ for the camera A, and $t_{H1} = 180$ for the camera B; $t_{H1} = 50$ for the camera A, and $t_{H1} = 210$ for the camera B; $t_{D1} = 30$; $t_{D1} = 150$; $t_N = 0.3$; $m = 19$; $p = 85$; $t_T = 0.4$. The number of ellipses used during the volume estimation is $n_v = 20$.

We collected a database of 20 subjects in uncontrolled light conditions, walking in 8 different directions, for a total of 160 pairs of frame sequences, with weights ranging from 43.7 kg to 101.1 kg. The weights are measured using a weighing scale

with an accuracy of ± 0.1 kg. The directions are chosen in order to cover all the possible situations of walking people (Fig. 4b), and uncontrolled light conditions permitted to better simulate real surveillance scenarios. The lengths of the frame sequences are different, according to the walking speed of the individuals. The minimum number of frames that compose a frame sequence is 7 and the maximum number is 29.

In order to search the best features, we created four feature datasets from the collected database:

- *Dataset 1*
  1) index of the walking direction (from 1 to 8);
  2) median of the volume vector V;
  3) value of the 10° percentile of V;
  4) value of the 90° percentile of V;
  5) median of the height vector H;
  6) value of the 10° percentile of H;
  7) value of the 90° percentile of H.
- *Dataset 2*
  1-7) Dataset 1;
  8-27) 20 values relative to the median of the areas of the 20 ellipses approximating the body shape (matrix A).
- *Dataset 3*
  1-7) Dataset 1;
  8-17) 10 values relative to the median of the areas of the ellipses approximating the torso ($n = 4 \dots 13$).
- *Dataset 4*
  1-17) Dataset 3;
  18-27) 10 values relative to the the 10° percentile of the areas of the ellipses approximating the torso ($n = 4 \dots 13$);
  28-37) 10 values relative to the the 90° percentile of the areas of the ellipses approximating the torso ($n = 4 \dots 13$) .

In order to compute the results depicted in the paper, a 10-fold cross-validation [26] was applied to every dataset separately. By using the cross-validation technique, we tested the ability of the neural networks to generalize, and adapt themselves to the environment and the acquisition scenario. In particular, the proposed method is based on feedforward neural networks composed by one input layer, one hidden layer and one output layer constituted by a linear node. The hidden layer was tested using different number of tan-sigmoidal nodes: in our experiments, we used 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 nodes. We used neural networks with a single hidden layer since they can be considered as universal approximators. All the neural networks were trained with the Levenberg-Marquardt back-propagation algorithm, using a maximum epoch limit equal to 150.

In order to prove the validity of the neural approach, we compared the results of the proposed method with the ones obtained by performing the weight approximation directly from the extracted silhouettes. The median of the volume of the individual related to every pair of frame sequences is estimated and then converted in kg by using approximating polynomials of different orders (from 1 to 20), computed using the least mean square technique. The best results obtained by the neural approach and by the reference method are summarized in Table I. It is possible to observe that our approach allows a more accurate weight estimation with respect to the direct computation, since it is less affected by problems related to the illumination condition, position, orientation, and speed of the walking person. The direct computation based on a polynomial approximation of the 10^{th} order, in fact, obtained a mean error near to 0 kg and a standard deviation of the error equal to 10.48 kg. Differently, the neural approach obtained a mean error equal to 0.07 kg and a standard deviation of the error equal to 2.30 kg.

TABLE I
RESULTS OF THE WEIGHT ESTIMATION ON THE EVALUATED DATASET

Method	Feature set	Mean error (kg)	Std error (kg)
FFNN-50	Dataset 1	0.18	6.05
FFNN-25	Dataset 2	0.07	2.30
FFNN-50	Dataset 3	-0.22	2.68
FFNN-30	Dataset 4	0.07	2.48
10^{th} order approx	Median of V	0.00	10.48

Notes: FFNN-x = feedforward neural network with one hidden layer composed by x nodes; 10^{th} order approx = direct approach based on the 10^{th} order polynomial approximation.

TABLE II
RESULTS OF THE WEIGHT ESTIMATION CONSIDERING DIFFERENT WALKING DIRECTIONS

Direction	Mean error (kg)	Std error(kg)
D1	0.01	0.31
D2	-1.13	4.77
D3	-0.14	0.65
D4	1.24	3.23
D5	0.11	0.39
D6	-0.05	1.74
D7	0.10	1.42
D8	0.40	1.55

Notes: results obtained on Dataset 2 using a feedforward neural network with one hidden layer composed by 25 nodes.

Table I also shows that the proposed neural approach is able to obtain the best results on Dataset 2. Consequently, for the considered database of frame sequences, the best evaluated feature set is composed by the walking direction, the characteristics related to the height and volume, and the median values of the areas of all the ellipses used for the approximation of the silhouette shape. This fact suggests that the use of information related to all the body parts permits to improve the accuracy of the weight estimation. However, in different applicative contexts, other features could obtain better results. For this reason, it is necessary to consider this aspect during the design and deployment of soft-biometric systems based on surveillance cameras.

In order to evaluate the robustness of the proposed method in the different walking directions, we evaluated the performance of the best trained neural network using the Dataset 2. The obtained results are shown in Table II. These results suggest that the proposed method can effectively estimate the weight of persons walking in different directions. The directions D2 and D4 obtained the highest mean error. This fact can be caused by less accurate silhouette segmentations due to the light conditions.

Lastly, we used the neural network that obtained the better accuracy and estimated the weight of every subject. The results

TABLE III
RESULTS OF THE WEIGHT ESTIMATION OF DIFFERENT INDIVIDUALS

Individual	True weight (kg)	Mean error (kg)	Std error(kg)
1	85.7	-0.64	1.57
2	93.3	-2.38	6.60
3	74.8	0.48	1.43
4	70.1	-0.09	0.40
5	81.3	1.53	3.82
6	54.2	0.62	1.55
7	76.6	0.68	2.04
8	100.5	0.12	1.08
9	80.8	-0.25	0.55
10	99.4	0.15	0.65
11	89.1	0.28	1.37
12	43.7	-0.05	1.05
13	67.9	1.25	3.48
14	70.6	-0.17	0.55
15	57.2	0.73	2.07
16	68.5	0.02	0.18
17	70.2	0.42	1.22
18	101.1	0.00	1.11
19	51.1	-0.36	1.06
20	77.4	-0.98	3.31

Notes: results obtained on Dataset 2 using a feedforward neural network with one hidden layer composed by 25 nodes.

are depicted in Table III. It is possible to observe that the obtained error is satisfactory for every considered individual. The maximum absolute mean error, in fact, is less than 2.4 kg.

V. CONCLUSIONS

In this paper, we proposed a method for a contactless, low-cost, unconstrained, and unobtrusive estimation of the weight of walking individuals in surveillance frame sequences. The weight estimation is performed using image processing techniques and a computational intelligence approach. For each pair of frame sequences, the human silhouette is segmented and a set of features are extracted. A neural approach is then used to process the extracted features, obtaining a weight estimation that is independent from the point of view, position, and illumination. Moreover, the method does not require the computation of complex models of the body parts.

We tested the proposed method on a database composed by frame sequences captured by two cameras in uncontrolled light conditions, describing eight different walking directions. The obtained results show that the method is feasible and can achieve an accurate weight estimation. The best mean error obtained on the evaluated set of frame sequences is equal to 0.07 kg, with a standard deviation of 2.30 kg. The obtained results also demonstrate that neural networks are effectively able to solve problems related to the variability of the environmental conditions. Moreover, the use of neural networks permitted to design a weight estimation method that is based on less complex features with respect to the other techniques in the literature that are able to perform a weight estimation from the visual aspect of the individuals.

REFERENCES

[1] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics (International Series on Biometrics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[2] A. K. Jain, S. C. Dass, K. Nandakumar, and K. N, "Soft biometric traits for personal recognition systems," in *International Conference on Biometric Authentication*, 2004, pp. 731–738.

[3] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-biometrics: unconstrained authentication in a surveillance environment," in *2009 Digital Image Computing: Techniques and Applications*, 2009, pp. 196–203.

[4] C. Velardo and J. Dugelay, "Weight estimation from visual body appearance," in *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6.

[5] B. Kamgar-Parsi, W. Lawson, and B. Kamgar-Parsi, "Toward development of a face recognition system for watchlist surveillance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1925–1937, October 2011.

[6] F. Wheeler, R. Weiss, and P. Tu, "Face recognition at a distance system for surveillance applications," in *IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–8.

[7] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *International Conference on Digital Image Computing: Techniques and Applications*, 2010, pp. 320–327.

[8] J. Zhang, Y. Cheng, and C. Chen, "Low resolution gait recognition with high frequency super resolution," in *PRICAI 2008: Trends in Artificial Intelligence*, ser. Lecture Notes in Computer Science, T.-B. Ho and Z.-H. Zhou, Eds. Springer Berlin / Heidelberg, 2008, vol. 5351, pp. 533–543.

[9] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in *Biometric Technology for Human Identification VII*, 2010.

[10] K. Niinuma, U. Park, and A. Jain, "Soft biometric traits for continuous user authentication," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 771–780, December 2010.

[11] S. Denman, A. Bialkowski, C. Fookes, and S. Sridharan, "Determining operational measures from multi-camera surveillance systems using soft biometrics," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2011, pp. 462–467.

[12] Y. Ran, G. Rosenbush, and Q. Zheng, "Computational approaches for real-time extraction of soft biometrics," in *19th International Conference on Pattern Recognition*, 2008, pp. 1–4.

[13] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 433–466, September 1995.

[14] H. Ando and H. Fujiyoshi, "Human-area segmentation by selecting similar silhouette images based on weak-classifier response," in *International Conference on Pattern Recognition*, August 2010, pp. 3444–3447.

[15] M. Piccardi, "Background subtraction techniques: a review," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, October 2004, pp. 3099–3104.

[16] Shireen, Khaled, and Sumaya, *Recent Patents on Computer Science*, vol. 1, pp. 32–34, 2008.

[17] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP J. Appl. Signal Process.*, pp. 2330–2340, January 2005.

[18] A. Sanin, C. Sanderson, and B. C. Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern Recognition*, vol. 45, no. 4, pp. 1684–1695, 2012.

[19] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *IEEE International Workshop on Information Forensics and Security*, 2010, pp. 1–6.

[20] R. Donida Labati, A. Genovese, V. Piuri, and F. Scotti, "Low-cost volume estimation by two-view acquisitions: A computational intelligence approach," in *International Joint Conference on Neural Networks*, 2012.

[21] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, November 2000.

[22] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *1997 Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106–1112.

[23] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1337–1342, 2003.

[24] A. Cavallaro and T. Ebrahimi, "Change detection based on color edges," in *IEEE International Symposium on Circuits and Systems*, vol. 2, May 2001, pp. 141–144.

[25] E. Jeges, I. Kispal, and Z. Hornak, "Measuring human height using calibrated cameras," in *2008 Conference on Human System Interactions*, 2008, pp. 755–760.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2nd ed. Wiley-Interscience, November 2001.