

Improving Person Recognition by Weight Adaptation of Soft Biometrics

Shaghayegh Gharghabi¹, Bitā Azari², Faraz Shamshirdar², Reza Safabakhsh²

¹Robotics Research Center, ²Department of Computer Engineering
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran

(s.gharghabi, bita.az, faraz.shamshirdar, safa)@aut.ac.ir

Abstract—One of the main challenges of current person recognition techniques lies on difficulties of recognition in various poses. Recently, attention has been focused on using soft biometric information extracted from the human body to overcome the biometric recognition system's limitation in unconstrained environments. In this paper, we integrate the face and body information in a linear combination. We propose a novel approach in which the weights of features in the recognition system are adapted based on the reliability of the detected joints extracted from the body and the correlation between features. We evaluate the proposed approach in recognizing a five person group in various poses such as sitting and circular walking. The method was applied to a service robot equipped with the Kinect sensor. The results show a mean improvement of 4.39% after weight adaptation based on the correlation between features and 6.88% after consideration of the reliability of the features.

Keywords—Weight Adaptation; Person Recognition; soft biometrics; Kinect.

I. INTRODUCTION

Automatic person recognition in various poses plays an important role in several fields including service robots, video surveillance and smart environments. Over the last years, a wide variety of biometric recognition systems (BRS) have been developed for person recognition. In these systems, traits such as face, gait, iris and fingerprint are used which can uniquely discriminate individuals [1]. Face is one of the most important biometrics for person recognition that has widely been studied in the past and reviewed in the literature [2, 3]. Face is a powerful and unique trait for person recognition; however it is vulnerable to pose, illumination variations and low resolution. Besides, in some cases, the face of a person might be occluded or invisible.

The research by Robbins et al. on humans [4] shows that the recognition rate of the people can be improved when both the body and face information are available and using such information is the most common method for recognition of the people. In addition to overcome the problems in primary recognition systems, recently a growing number of researches have focused on exploiting ancillary information based on the description of human physical features known as soft biometric traits [1]. The global shape of the body and the skeleton data

can be exploited in unconstrained environments in a non-intrusive way and make it possible to distinguish between people when face cue is not available or not sufficient for obtaining accurate results in the long-time span.

A few recent works have proposed to use the combination of the human physical features and face in recognition systems. The combination of these biometrics based on 2D images has been suggested in [5]. This work showed how soft biometrics can improve the recognition accuracy of the face recognition system at a distance. In [6] the cues of a poselet-level person recognizer with a face recognizer and a global recognizer for person recognition in photo albums are accumulated. In this method, it is shown that the combination of these recognizers improves the performance of person recognition. The recognition based on 2D images has some drawbacks, they estimate the soft biometrics of a person based on the sequence of frames or an annotator determines the score of soft biometrics. Thus these methods are not as accurate as 3D approaches.

In recent years, the appearance of cheap 3D sensors such as Kinect has facilitated the acquisition of 3D information of skeleton data such as the height and limb lengths. There are a few works focusing on recognition based on the depth information for extracting skeleton features. In a research by Meifen Cao et al. [7] PCA is used to extract an individual's body features for recognizing people. This method applies PCA to 19 link lengths of the individual. Then the feature vector is obtained and used for the recognition and tracking of the individual by the robot. The use of head and shoulders area in 3D range data for recognizing individuals is proposed in [8]. In this approach, person recognition is performed based on a support vector machine (SVM) in which the head to shoulder signature (HSS) is utilized as the basis. In [9] skeleton-based and surface-based features are utilized for recognition. They weight the features by using quasi-exhaustive learning strategy to maximize the re-identification performance. The other research by Matteo et al. used the body joints as keypoints for extracting 3D descriptors for re-identification. In this method, the reliabilities of extracted joints are evaluated to consider joints as the keypoints [10]. In our previous work we combine the face and body information for non-invasive automatic person recognition in various poses and low resolution [11].

In this research, as the body features show different discriminative strengths, we assign weights to the features.

These weights are based on the correlation between features to improve the person recognition accuracy. In addition, the extracted skeleton features may have low reliability in various poses. Then we propose a method to adapt the weights of features based on their reliability. The rest of this paper is organized as follows. The proposed person recognition system and its components are described in detail in Section II. Experimental setup and results are presented in Section 0. Finally, conclusions and proposal for future work proposed are given in Section IV.

II. PERSON RECOGNITION SYSTEM

In this section we combine face and body recognition systems. In face recognition system the sparse classifier is used to obtain the probability of belonging the test face to each training class. Moreover, some body features have been used in the person recognition system. These features were weighted using the correlation between features based on the reliability of features extracted from the body. We describe our system in detail in the following sub-sections.

A. Computing and combination of classifiers

We used the sparse representation classification (SRC) technique proposed by Wright et al. [12] for face recognition. The main idea in this method is that sparse nonzero coefficients concentrate on training samples of the same class of the test sample. A test sample coded as a sparse linear combination of the training images and its 1-norm minimization is calculated. The SRC method classifies the test sample by evaluating which class of training samples has the minimum reconstruction error with the associated coding coefficients. This method is selected because of its high performance reported in former studies [13], in addition to being real-time. The SRC is applied on the detected faces by the well-known Viola-Jones [14] method.

We also automatically extract the soft biometric characteristic from the information of the individual's body by using OpenNI skeleton tracker which is commonly used in computer vision and robotics [15]. OpenNI tracker works on 3D data and provides the coordinate of 3D positions for 15 different skeletal joints of a person in real time. This tracker has a superior performance over the 2D-based approaches for estimating the coordinate of joints positions [16]. We extract the 14 skeleton features from the coordinate values of the detected joints (Fig. 1). These features are the distances between the joints which are calculated using the Euclidean distance. We considered 15 classifiers for face and 14 body features.

We considered c as the label of individual's identity class and F as the concatenation of the individual's face and body features. The identity of an individual is predicted by using a linear combination of the predicted probabilities of classifiers as:

$$s(F, c) = \sum_i \omega_i P_i(c|F) \quad (1)$$

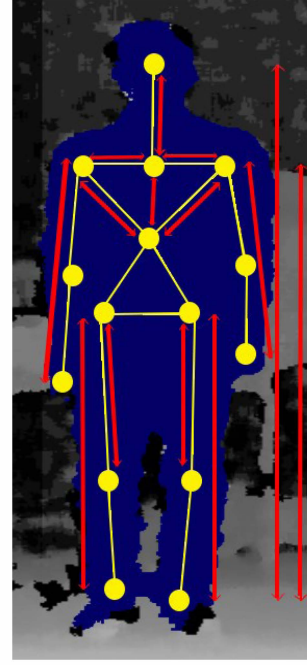


Fig. 1. Link length between adjacent joints and the selected features from body information

Here $P_i(c|F)$ is the probability of identity class c given by part i for feature F and ω_i is the associated weight for the classifier of the i th part. By using sparse face recognition each identity class c is predicted. We refer to its predication as $P_0(c|F)$. The distributions for the skeleton features are considered Gaussian. Therefore, the $P_i(c|F), i=1, \dots, 14$ is calculated using the probability density of the Gaussian distribution. In the next section, we will describe the method we use to compute ω_i .

A. Correlation

The features in the proposed method are weighted. These weights determine the role of features for recognition of individuals. To understand these roles, the internal correlation and discriminative power of all the features are calculated. The correlation between all pairs of features is computed using the Pearson's correlation coefficient:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{j=1}^N (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^N (X_j - \bar{X})^2} \sqrt{\sum_{j=1}^N (Y_j - \bar{Y})^2}} \quad (2)$$

where σ_X and σ_Y are the standard deviations of the two variables X and Y , also σ_{XY} represents their covariance. X_j and Y_j are the two distinctive features of the j th person and the $\bar{X}$ and $\bar{Y}$ are the average of those features for all individuals.

The values of r range are from -1.0 to 1.0 . The correlation coefficients close to 1.0 imply that a linear equation perfectly

describes the relationship between X and Y , with all data points lying on a line for which Y increases as X increases. A value of -1.0 implies that all data points lie on a line for which Y decreases as X increases. There is no linear correlation between the variables when the value of correlation coefficient is 0.

The correlation matrix of the features is obtained by Equation (2). Summing up the rows of the matrix, we calculated the correlation between each feature and all other features. Since the information of the extracted body features with a greater correlation value is the same as the other features we assigned a lower weight to that feature. In contrast, a feature with a lower correlation value should have a higher weight. To satisfy these goals, we represented a correlation-based weight adaptation (CWA) algorithm. In this approach, the average of each rows' numbers are used to find the correlation of the feature in comparison with others. The obtained correlations are $a_i, i = 1, \dots, 14$. Then the weights of all parts are calculated based on the maximum ($a_{\max}$) and the minimum ($a_{\min}$) of the average correlations by:

$$w_i = \frac{a_{\max} + a_{\min} - a_i}{n(a_{\max} + a_{\min}) - as} \quad (3)$$

where n is the number of features and as is the summation of all average correlation values.

B. Reliability

In some poses, the skeletal tracker may not be able to detect all the joints possibly caused by the quality of the segmented target and the level of occlusion occurred in different poses. In this case, OpenNI tracker marks those joints as untracked. Thus, the locations of the joints are estimated, but they are not as reliable as the locations of the other joints. We proposed a method in which we applied the reliability-based weight adaptation to the CWA method and named it RCWA.

Since all the features are parts of the whole skeleton, although the joints are not reliable, we extracted the features from them and changed the weights of the features based on their reliabilities. Two versions of the weight adaptation algorithm were developed: 1) we considered the feature completely unreliable, thus a zero value is assigned as the weight of the feature obtained from untracked joints. Therefore in this case just the reliable features (RF) participate for person recognition. 2) we assigned weights to unreliable estimated joints. These weights are calculated based on the number of the reliable joints. All the features are participating for recognition (AF), but their weights are adapted based on the number of reliable joints using:

$$\omega'_i = \omega_i \times \frac{\text{number of reliable joints}}{\text{number of all joints}} \quad (4)$$

This weight adaptation method is based on the assumption that if the number of reliable joints is greater, then the estimated joints are more reliable. Thus the weights are updated along the time based on the reliability of the joints.

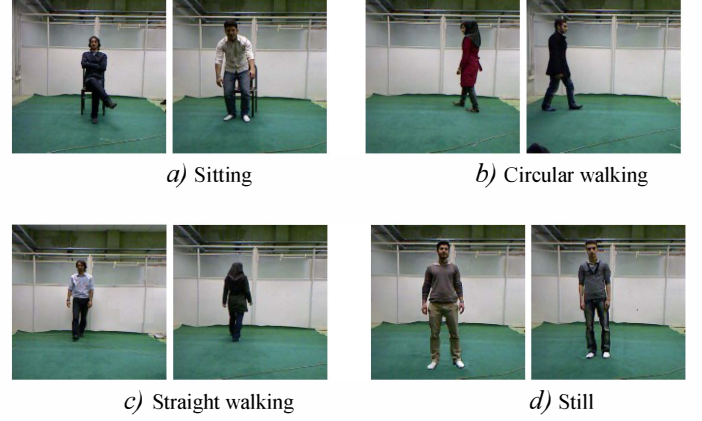


Fig. 2 Sample images of persons in four different poses that were used for training and recognition.

III. EXPERIMENTS

This section presents results from an empirical evaluation of the proposed person recognition system. As there was no dataset for person recognition in different poses containing the 3D information of the joints, we prepared a dataset. We used this application for domestic robots which should recognize several individuals in the environment. Hence the proposed method is evaluated only on a five person group; three adult men and two adult women. In this experiment, we focused on recognizing persons in different poses. In every pose, only some of the body information or face is available and the reliability of the individuals' joints can differ from pose to pose. The data for evaluation was collected in different days.

We considered four different sequences of straight walking, circular walking, sitting, and standing still. Standing still is used for training and the other three sequences are used as our test set. In the sitting sequence, both the body and the face information are available, but the skeleton information cannot be determined with high accuracy, resulting in low reliability for the joints. In addition the face resolution is low for recognition. In straight walking, when the person walks towards the camera, both the face and the body are visible, but when the person walks away, the face is not visible and only the body information is accessible. When the person walks in a circular path, some joints of the body are not visible and the face is visible from the side view. Some of sample images of persons in different poses are shown in Fig. 2.

In the still pose, each person stands in front of the camera for five seconds and the data are saved into a dataset to be compared with the test sample. For each person, 5×30 frames will be available. From every 15 frames, an image of the face is saved, resulting in ten images of the face for every person. The data for each test sample sequence is collected from the moment the person appears in the scene until 20 seconds later. For each person, the skeleton features are computed. Then by taking the average of them, a vector of 14 skeleton features is formed and saved into the dataset for each person.

Logically, we assumed the correlation between the face and the body features is zero. The correlation matrix containing the correlation between all skeleton features is represented

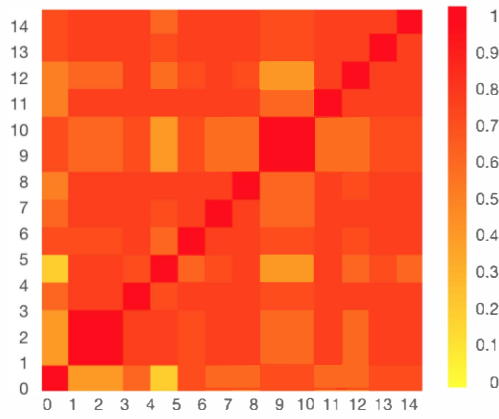


Fig. 3 The correlation matrix between body features

graphically in Fig. 3. All the features had positive correlations, shown with a range of red colors. Correlation coefficients close to 1.0 are represented in deep red and correlation coefficients close to 0 are represented in pale red. In this dataset, the accuracy of the proposed weight adaptation system compared to some standard other methods is presented. We reported the rates for the face system, body system, and the combination of them as the baselines. In the combination of face and body, we considered equal weights for the skeleton features without using the face information. In this case, the weight for the face recognition system is considered 0.7 and the remaining weights are assigned a value of 0.3. In addition, the accuracy of the proposed methods for weight adaptation (CWA, RCWA^{RF}, RCWA^{AF}) are reported.

Table I lists the recognition rates for four testing sets by the competing methods. By taking the average of all the sample sequences, we achieved the accuracy of 40.48% for the face recognition system, 67.23% for the body recognition system and 80.23% for the combination of them. The CWA system improved the person recognition rate by 4.39% compared to the combination of face and body without weight adaptation. In addition, by considering the reliability of features for combining them in RCWA^{RF}, we obtained a higher performance. Finally since unreliable features contain some useful information for person recognition, RCWA^{AF} achieved the best performance in comparison to other methods.

TABLE I. RECOGNITION RATES OF FACE AND BODY RECOGNITION SYSTEM, THEIR COMBINATION AND OUR PROPOSED METHODS

	Sitting(%)	Circular walking(%)	Walking straight(%)	Average(%)
Face	76.84	10.19	34.42	40.48
Body	44.93	71.61	85.15	67.23
Face-body	81.69	72.33	86.69	80.23
CWA	90.96	75.28	87.63	84.62
RCWA^{RF}	92.32	76.36	87.73	85.47
RCWA^{AF}	95.01	78.21	88.13	87.11

IV. CONCLUSION

This paper presented a method in which the body features are utilized and combined with the face information for solving the problems of person recognition in different poses. Integration of face and body biometrics was performed using a linear combination. Our specific aim was to improve person recognition accuracy by adaptation of weights for soft biometrics based on the correlation between features and the reliability of detected joints. It is shown that the proper association weights for features can improve the recognition accuracy. We also demonstrated that assigning lower coefficients to the features with higher correlations leads to a higher accuracy in establishing the user identity. This method is used on a service robot by utilizing RGBD sensor (Microsoft Kinect).

In future work, we will study recognition of partly occluded bodies and bodies in more varied poses. In addition, since the appearance of people changes with time (as an example growing mustache and their global shape of the body), we plan to develop an unsupervised algorithm to adapt the training set and learn the new features for persons along the time.

REFERENCES

- [1] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*, pp. 731-738. Springer Berlin Heidelberg, 2004.
- [2] J. H. Shah, M. Sharif, M. Raza, and A. Azeem, "A survey: Linear and nonlinear pca based face recognition techniques," in *International Arab Journal of Information Technology*, Vol. 10, No. 6, pp. 536-545, 2013.
- [3] T. Sinsinwar, and P. K. Dwivedi, "A survey on feature extraction method and dimension reduction techniques in face recognition," in *International Journal of Applied Science and Engineering Research*, Vol. 3, No. 3, pp. 734-740, 2014.
- [4] R. A. Robbins, and M. Coltheart, "The effects of inversion and familiarity on face versus body cues to person recognition," in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 38, No. 5, pp. 1098-1104, 2012.
- [5] Tome, Pedro, et al. "Soft biometrics and their application in person recognition at a distance," in *IEEE Transactions on Information Forensics and Security*, pp. 464-475, 2014.
- [6] Zhang, Ning, et al. "Beyond Frontal Faces: Improving Person Recognition Using Multiple Cues." arXiv preprint arXiv:1501.05703, 2015.
- [7] M. Cao, and H. Hashimoto, "Specific person recognition and tracking of mobile robot with Kinect 3D sensor," in *39th IEEE Annual Conference on the Industrial Electronics Society (IECON)*, pp. 8323-8328, November 2013.
- [8] N. Kirchner, A. Alempijevic, and A. Virgona, "Head-to-shoulder signature for person recognition," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1226-1231, 2012.
- [9] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with rgb-d sensors," in *Computer Vision-ECCV Workshops and Demonstrations*, pp. 433-442. Springer Berlin Heidelberg, 2012.
- [10] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, "A feature-based approach to people re-identification using skeleton keypoints," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.
- [11] S. Gharghabi, and R. Safabakhsh, "Person recognition based on face and body information for domestic service robots," in *IEEE 3rd RSI International Conference on Robotics and Mechatronics (ICROM)*, pp. 265-270, 2015.

- [12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210-227, 2009.
- [13] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [14] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. I-511, 2001.
- [15] OpenNI, OpenNI. <http://openni.org/> (Last Accessed: February 2015).
- [16] M. Raptis, D. Kirovski, and H. Hoppe, "Real-time classification of dance gestures from skeleton animation," in *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation*, pp. 147-156, 2011.