SHOW ME YOUR BODY: GENDER CLASSIFICATION FROM STILL IMAGES

Ioannis A. Kakadiaris[1] *Nikolaos Sarafianos*[1] *Christophoros Nikou*[1,2]

[1] Computational Biomedicine Lab, Dept. of Computer Science, Univ. of Houston, Houston, TX
[2]Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece

ABSTRACT

In this work, we investigate the problem of predicting gender from still images using human metrology. Since the values of the anthropometric measurements are difficult to be estimated accurately from state-of-the-art computer vision algorithms, ratios of anthropometric measurements were used as features. Additionally, since several measurements will not be available at test time in a real-life scenario, we opted for the Learning Using Privileged Information (LUPI) paradigm. During training, we used as features, ratios from all the available anthropometric measurements, whereas at test time only ratios of measurable (i.e., observable) quantities were used. We show that by using the LUPI framework, the estimation of soft biometric characteristics such as gender is possible. Gender classification from human metrology is also tested on real images with promising results.

***Index Terms*—** Gender Classification, Privileged Information, Anthropometry, Soft Biometrics

1. INTRODUCTION

Identifying humans based on soft biometrics has been an active area of research over the past decade [1]. In the work of Jain *et al.* [2], a framework for integrating soft biometric information such as gender, height, weight, age, and ethnicity with the output of the primary biometric system (e.g., face) is introduced. Most of the existing work [3, 4, 5] has approached the problem of classifying humans based on soft biometrics originating from facial information. However, in a real-life surveillance scenario, information from the face is not always available, as the face might be covered or occluded. Hence, approaches that employ information from the whole body to identify soft biometrics such as gender were introduced [6, 7, 8]. Anthropometric measurements can be exploited as soft biometrics, since they can be obtained at a distance or from a single image and they can also be employed to reduce the search space [9]. Even if the body is partly visible (i.e., occluded) and gait information cannot be exploited, several such anthropometric ratios can be extracted and contribute to human identification. In the work of Adjeroh *et al.* [10], the correlation of a variety of anthropometric features from the CAESAR anthropometric database [11] was investigated and a cluster-driven prediction model based on human metrology was introduced. Using the same dataset, Cao *et al.* [12] proposed a copula model (i.e., the marginal probability distribution of each variable is uniform) to predict the gender and the weight from human metrology. Finally, the use of ratios of anthropometric measurements has also been employed for person identification from Kinect videos in the work of Munsell *et al.* [13].

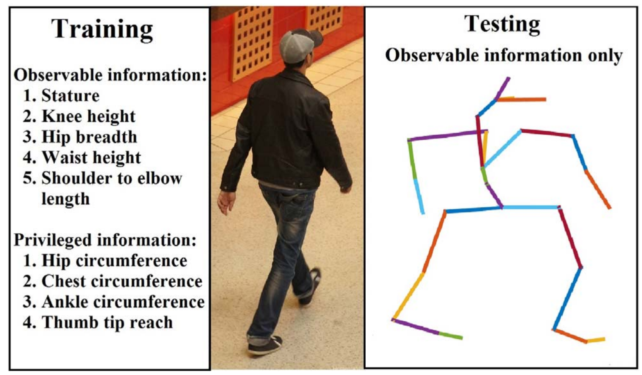

Fig. 1: Employing privileged information in human metrology for gender prediction. Additional information is provided about the anthropometry of the human model only during training. Photo by Hotlanta Voyeur is licensed under CC BY.

The Learning Using Privileged Information (LUPI) framework, which is implemented using the SVM+ algorithm, was introduced by Vapnik [14], and Vapnik and Vashist [15]. The new learning model places a nontrivial teacher in the training process who supplies the training set with additional information (i.e., features) that is not available for test examples. Following that, fast algorithms for solving the optimization problem of SVM+ were proposed [16], new approaches [17, 18, 19, 20] were introduce dWand discussions [21] on the LUPI paradigm have been ensued. Only recently, the LUPI paradigm was employed with applications to biometrics such as face verification and person re-identification [22], and soft biometrics for age estimation [23].

We propose a novel method, which performs gender (binary) classification using ratios of anthropometric measurements. Figure 1 depicts an overview of our approach when the LUPI [15] paradigm is employed. Our work follows a similar approach with the method of Cao *et al.* [12], in the sense that the same anthropometric database is used to predict soft biometric attributes. However, there are two distinctive differences. First, since the actual anthropometric measurements are highly unlikely to be obtained accurately from state-of-the-art computer vision algorithms that use images or videos captured from surveillance cameras, we opted for using ratios of anthropometric measurements. Second, we argue that several anthropometric measurements are relatively difficult to be estimated automatically (e.g., circumferences of human parts) and that such information will not be available in automatically acquired data. To the best of our knowledge, this the first time that: (i) ratios of anthropometric measurements are used in a scenario that does not employ depth information, and (ii) human metrology is used to predict soft biometric attributes from real images.

2. GENDER PREDICTION WITH LUPI

Below, we introduce the sets of features that will be used for each task, review the concepts of SVM for completeness, and describe the SVM+ algorithm and the Margin Transfer method of Sharmanska *et al.* [18]. Throughout this section the same notation as in the work of Sharmanska *et al.* [18] is used.

Ratios of anthropometric measurements: Based on the findings of the work of Cao *et al.* [12], using the actual values of anthropometric measurements (e.g., limb lengths in *mm*) from an anthropometric database results in good gender classification accuracy. We argue though, that such information cannot be accurately obtained from state-of-the-art computer vision algorithms without employing depth information (e.g., use data obtained from a Kinect RGB-D sensor). To address this limitation, we propose to exploit the use of ratios of anthropometric measurements. Hence, errors during the estimation of the actual values will be alleviated. A variety of anthropometric measurements from the body and the head is provided in the CAESAR database [11]. We divide these measurements into two groups. The first group contains only ratios of body measurements that can be captured from a regular surveillance camera and computed from state-of-the-art computer vision algorithms. This set, which will be denoted as X, contains only observable information (e.g., arm or leg lengths) and will be available during both the training and the testing phases. The second group, which will be denoted by X^*, contains ratios of body measurements that are difficult to obtain with an automated acquisition system (e.g., circumferences of body parts) as well as a few measurements that correspond to the head (e.g., head breadth or face length). This type of information is considered as privileged and it will not be available at test time.

From SVM to SVM+: In the standard paradigm of supervised learning for binary classification, the training set consists of N pairs of feature vectors x_i, along with their respective labels y_i, represented as $(x_1, y_1), \ldots, (x_N, y_N)$, $x_i \in \mathbb{R}^d$ where d is the number of features of each sample and $y_i \in \{-1, +1\}$. The standard SVM classifier finds a maximum-margin separating hyperplane between the two classes and solves the following constrained optimization problem:

$$\begin{aligned} \underset{\substack{\xi_1,\ldots,\xi_N,\\ w,\, b}}{\text{minimize}} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i, \\ \text{subject to:} \quad & y_i\big(\langle w, x_i\rangle + b\big) \geq 1 - \xi_i \\ & \xi_i \geq 0,\ i = 1, \ldots, N \end{aligned} \tag{1}$$

where $w \in \mathbb{R}^m$ represents the weight vector, $||w||^2$ indicates the size of the margin, $b \in \mathbb{R}$ is the bias parameter, ξ is the slack variable for one training sample and indicates the deviation from the margin borders and C denotes the penalty parameter. In a LUPI setup, during the training phase, instead of pairs we have triplets $(x_i, x_i^*, y_i), x \in \mathbb{R}^d, x^* \in \mathbb{R}^{d^*}$, $y_i \in \{-1, +1\}$, where feature vectors x^* represent the additional (i.e., privileged) information. During the testing phase, features from the privileged space X^* are not available. The goal of LUPI is to exploit the privileged information during the training phase to learn a model that further constrains the solution in the original space X and thus, it can describe more accurately the testing data. In this paradigm, the slack variables ξ_i are parameterized as a linear function of privileged information $\xi_i(w^*, b^*) = \langle w^*, x_i^*\rangle + b^*$. The SVM+ problem in the training phase solves the following minimization problem:

$$\begin{aligned} \underset{w,\, b,\, w^*, b^*}{\text{minimize}} \ & \frac{1}{2}\big(||w||^2 + \gamma||w^*||^2\big) + C\sum_{i=1}^{N}\xi_i(w^*, b^*). \\ \text{subject to:} \ & y_i\big(\langle w, x_i\rangle + b\big) \geq 1 - \xi_i(w^*, b^*) \\ & \xi_i(w^*, b^*) \geq 0,\ i = 1, \ldots, N \end{aligned} \tag{2}$$

Margin Transfer: Sharmanska *et al.* [18] investigated the framework of using privileged information for object recognition and introduced a Margin Transfer approach. The proposed method interprets the LUPI concept as learning the *easiness* and *hardness* of each sample to be classified based on the margin distance to the classifying hyperplane in the privileged space. This knowledge is then transferred to the original space to train a classifier with improved performance. A standard SVM classifier is first trained on X^*, the prediction function of which is $f^*(x^*) = \langle w^*, x^*\rangle$, and the margin distance $\rho_i := y_i f^*(x_i^*)$ between the training samples and the decision function in the privileged space is computed. Large values of ρ_i indicate that the respective sample can be classified easily, low values correspond to samples that are more difficult to classify, and negative values samples that are impossible to classify. The minimization problem is formulated as follows:

$$\begin{aligned} \underset{w \in \mathbb{R}^d,\ \xi_i \in \mathbb{R}}{\text{minimize}} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i. \\ \text{subject to:} \quad & y_i\langle w, x_i\rangle \geq \rho_i - \xi_i \\ & \xi_i \geq 0,\ i = 1, \dots, N \end{aligned} \tag{3}$$

Unlike SVM+, the performance of the classifier in the privileged space is very important for the Margin Transfer method because the information in the privileged space defines the margin in the original space.

3. EXPERIMENTAL EVALUATION

For the purpose of this paper, we used the CAESAR database [11] which is comprised of 44 anthropometric measurements (in *mm*), the weight (in *kg*) and the gender of 2,392 US and Canadian civilians. After data-preprocessing and discarding data with missing values, the size of the dataset used for the experimental evaluation is 2,369×39 including the gender. The ratios of anthropometric measurements we obtained are split to: (i) X which contains $(12\times11)/2 = 66$ observable features (i.e., ratios) for each human subject and (ii) the privileged set X^* with size of $(26\times25)/2 = 325$ for each sample. Furthermore, since in many real-life scenarios some of the observable features (X) may not be available (e.g., due to occlusions) we investigated separately the cases of having, at test time, features only from the upper body (e.g., arm lengths) or only from the lower part of the human body (e.g., knee height). When no privileged information is used, we opted for a linear SVM which requires only the penalty parameter C to be cross-validated. For SVM+, we used a linear kernel in the original space and a radial basis function kernel type in the correction space. In the latter case, additional tuning of the kernel coefficient γ in the correcting space is necessary. The search-space for both C and γ were $[10^{-4}, 10^{-3}, \dots, 10^{4}]$. A standard 5-fold cross-validation scheme was selected and a full-grid search was performed.

Gender Classification on the CAESAR dataset: In Table 1, we present the results of the proposed approach using the standard SVM and SVM+ methods when the testing set comprises observable features: (i) from the whole human body (i.e., X), (ii) only the lower body (i.e., X_L), and (iii) only the upper body (i.e., X_U). The last two rows correspond to classification results when all the measurements are used in both training and testing and can be interpreted as an upper boundary for the classification performance. The second to last row uses ratios of anthropometric measurements as features, whereas the method of Cao *et al.* [12] uses the corresponding actual values of the measurements. When all the observable features are used, the LUPI paradigm improves gender classification accuracy. Interestingly, features from the lower part of the body exhibited a significantly better classification accuracy compared to the ones obtained from the upper body. When only upper body features are available at test time, the LUPI framework does not appear to result in increased accuracy, and in both methods the classification accuracy is reduced to 76.6 % while the standard deviation increases to almost 3 %.

Testing Features	SVM	SVM+
X	97.61 ± 0.44	**98.18 ± 0.56**
X_L	95.34 ± 0.74	**95.82 ± 0.81**
X_U	**76.69 ± 2.98**	76.54 ± 2.95
$X \cup X^*$	99.10 ± 0.23	-
Cao *et al.* [12]	99.37	-

Table 1: Gender classification mean accuracy (%) and standard deviation on the CAESAR dataset using SVM and SVM+. The first column denotes which of the observable features are available at test time. The last two rows employ all the available information during both training and testing, and thus a LUPI framework is not applicable.

Leveraging Privileged Information in Human Metrology: Beneficial or Redundant? Based on the results obtained from the SVM+ method, the LUPI framework can improve the classification accuracy. However, two important characteristics of LUPI have to be taken into consideration. The first is what may be considered privileged information. Although conceptually it might be appealing to use circumferences of human limbs as prior information to boost the gender prediction accuracy, we demonstrated that this is not always true (e.g., when only upper body features are visible). The second is how the existence of privileged information is exploited to improve the accuracy during test time. To investigate this issue, we used the Margin Transfer algorithm of Sharmanska *et al.* [18] and report gender classification accuracy results in Figure 2. In contrast to the SVM+ approach, in the Margin Transfer method the performance of the classifier in the privileged space is of significant importance, and thus it is denoted with *SVM - Priv. Space*. Interpreting the LUPI framework in this manner results in a slightly worse classification accuracy compared to a standard SVM. The reason is that the performance in the privileged space is worse than in the original data.

Gender Classification on real data: Our next objective is to use the discriminative power of the anthropometric measurements to perform gender classification on humans in real images. We opted for the Point and Shoot Face Challenge (PaSC) dataset [24] which is comprised of 2,802 videos of 293 people who perform a simple action while walking towards the camera and the SARC3D dataset introduced by Baltieri *et al.* [25]. For each human in the PaSC dataset, we selected the first frame of the respective video in order for the full body to be visible and created a dataset with 248 images of 116 men and 132 women. Since in the PaSC images the

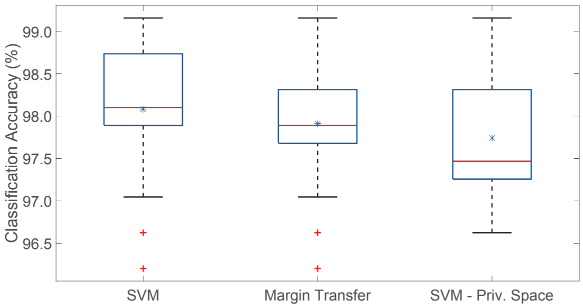

Fig. 2: Gender classification accuracy (%) in 20 random train-test splits using a standard SVM, the Margin Transfer method [18], and a standard SVM in the privileged space.

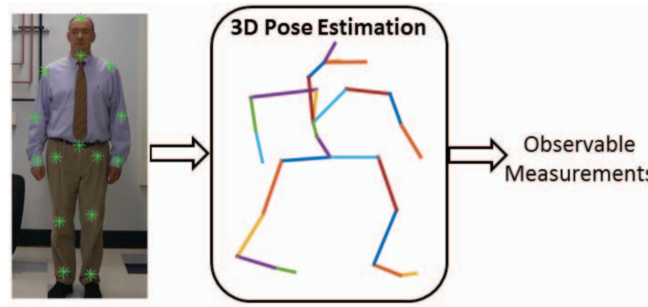

Fig. 3: Depiction of extracting anthropometric features from images. Given an image with 2D annotations, 3D pose estimation is performed to obtain the landmarks in 3D and compute the ratios of observable measurements

face is visible and gender prediction can be performed using facial information, we also selected all 50 images from the SARC3D dataset, which depicts people walking away from the camera. The number of men and women in SARC3D is 43 and 7, respectively. In both datasets, we annotated 15 main joints in 2D in order to avoid an additional source of error originating from the 2D pose estimation algorithm. For a completely automated system, a state-of-the-art 2D pose estimation algorithm [26] can be employed. We then used the method of Zhou *et al.* [27] to perform 3D pose estimation for each human and computed their observable features (i.e., ratios of limb lengths). A representative example of the feature extraction process is shown in Figure 3. Gender classification results are reported in Table 2 using a standard SVM for 20 random train-test splits. The reason the reported classification accuracy on SARC3D is higher than PaSC, can be attributed to the fact that the 3D poses in SARC3D are easier to be estimated since people are standing in an upright position. Columns two to four in Table 2 contain different results depending on which parts of the human body are visible. Identifying the gender of humans from real images using anthropometric measurements is a challenging task. Its difficulty arises from the fact that the 3D pose estimation algorithm starts with an initial 3D pose and a dictionary of poses (i.e., bases) and by exploiting this information maps the 2D joint locations to 3D through an optimization scheme. The performance of this algorithm is sensitive to the initialization, and

	Set of features		
Dataset	X	X_L	X_U
PaSC	71.37 ± 1.64	57.65 ± 2.82	58.06 ± 2.73
SARC3D	86.00 ± 2.00	78.00 ± 4.00	72.00 ± 4.00

Table 2: Gender classification accuracy (%) on still images from the PaSC and SARC3D datasets using an SVM when the feature set contains full, upper and lower-body features.

thus the performance is not always robust. Note that employing privileged information from the CAESAR dataset, and exploiting this information at test time using as features the obtained measurements from the images, resulted in a worse performance. Thus, the LUPI paradigm using anthropometric measurements estimated from images was not investigated further.

4. CONCLUSION

In this paper, we proposed a novel method of performing gender classification using human metrology. Feature vectors comprise ratios of anthropometric measurements because they can be estimated more accurately than actual measurements in *mm* from state-of-the-art computer vision techniques. Most of the features available during training will not be available at test time in real-life scenarios. Consequently, we opted for the Learning Using Privileged Information framework. We observed that using privileged information usually results in more accurate gender classification than the standard SVM. Results are reported under different scenarios where part of the body is not fully visible. Using real images from the PaSC and SARC3D datasets, we used a state-of-the-art 3D pose estimation algorithm to obtain the joint locations in three dimensions, computed the ratios of the respective limb lengths, and presented promising results.

Acknowledgments

This work has been funded in part by the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. The work of C. Nikou is supported by the European Commission (H2020-MSCA-IF-2014), under grant agreement No 656094. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

5. REFERENCES

[1] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 441–467, March 2016. 1

[2] A. K Jain, S. C Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Proc. International Conference on Biometric Authentication*, Hong Kong, China, July, 15-17 2004, pp. 731–738. 1

[3] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 541–547, 2008. 1

[4] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, 2012. 1

[5] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K Pal, C. Bhagavatula, and M. Savvides, "Investigating the feasibility of image-based nose biometrics," in *Proc. IEEE International Conference on Image Processing*, Quebec City, Canada, Sep. 27-30 2015, pp. 522–526. 1

[6] L. Cao, M. Dikmen, Y. Fu, and T. S Huang, "Gender recognition from body," in *Proc.* 16^{th} *ACM International Conference on Multimedia*, Vancouver, British Columbia, Canada, Oct. 27-31 2008, pp. 725–728. 1

[7] G. Guo, G. Mu, and Y. Fu, "Gender from body: A biologically-inspired approach with manifold learning," in *Proc. Asian Conference on Computer Vision*, 2010, pp. 236–245. 1

[8] D. Freire-Obregon, M. Castrillon-Santana, E. Ramon-Balmaseda, and J. Lorenzo-Navarro, "Automatic clothes segmentation for soft biometrics," in *Proc. IEEE International Conference on Image Processing*, Paris, France, Oct. 27-30 2014, pp. 4972–4976. 1

[9] Antitza Dantcheva, Arun Singh, Petros Elia, and Jean-Luc Dugelay, "Search pruning in video surveillance systems: Efficiency-reliability tradeoff," in *Proc. IEEE International Conference on Computer Vision Workshops*, Barcelona, Spain, Nov. 6-13 2011, pp. 1356–1363. 1

[10] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *Proc. IEEE International Workshop on Information Forensics and Security*, Seattle, WA, Dec. 12-15 2010. 1

[11] SAE International, "CAESAR: Civilian American and European Surface Anthropometry Resource database," (Available online at http://store.sae.org.caesar). 1, 2, 3

[12] D. Cao, C. Chen, D. Adjeroh, and A. Ross, "Predicting gender and weight from human metrology using a copula model," in *Proc.* 5^{th} *IEEE International Conference on Biometrics Theory, Applications and Systems*, Washington DC, USA, Sep. 23 - 26 2012, pp. 162–169. 1, 2, 3

[13] B. C Munsell, A. Temlyakov, C. Qu, and S. Wang, "Person identification using full-body motion and anthropometric biometrics from kinect videos," in *Proc. European Conference on Computer Vision. Workshops and Demonstrations*, Florence, Italy, Oct. 7-13 2012, pp. 91–100. 1

[14] V. Vapnik, *Estimation of dependences based on empirical data*, Springer Science & Business Media, 2006. 1

[15] V. Vapnik and A. Vashist, "A new learning paradigm: Learning using privileged information.," *Neural Networks*, vol. 22, no. 5-6, pp. 544–57, 2009. 1

[16] D. Pechyony, R. Izmailov, A. Vashist, and V. Vapnik, "SMO-style algorithms for learning using privileged information," in *Proc.* 6^{th} *International Conference on Data Mining*, Las Vegas, Nevada, July 12-15 2010, pp. 235–241. 1

[17] V. Sharmanska, N. Quadrianto, and C. H Lampert, "Learning to rank using privileged information," in *Proc. IEEE International Conference on Computer Vision*, Sydney, Australia, Dec. 3-6 2013, pp. 825–832. 1

[18] V. Sharmanska, N. Quadrianto, and C. H Lampert, "Learning to transfer privileged information," *arXiv preprint arXiv:1410.0389*, 2014. 1, 2, 3, 4

[19] M. Lapin, M. Hein, and B. Schiele, "Learning using privileged information: SVM+ and weighted SVM," *Neural Networks*, vol. 53, pp. 95–108, 2014. 1

[20] L. Liang and V. Cherkassky, "Connection between SVM+ and multi-task learning," in *Proc. IEEE International Joint Conference on Neural Networks*, Hong Kong, June 1-8 2008, pp. 2048–2054. 1

[21] C. Serra-Toro, V J. Traver, and F. Pla, "Exploring some practical issues of SVM+: Is really privileged information that helps?," *Pattern Recognition Letters*, vol. 42, pp. 40–46, 2014. 1

[22] X. Xu, W. Li, and D. Xu, "Distance metric learning using privileged information for face verification and person re-identification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. pp, no. 99, pp. 1, 2015. 1

[23] S. Wang, D. Tao, and J. Yang, "Relative attribute SVM+ learning for age estimation," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 827–839, 2015. 1

[24] J R. Beveridge, P J. Phillips, D. Bolme, B. A Draper, G. H Givens, Y. M. Lui, M. N. Teli, H. Zhang, W T. Scruggs, K. W Bowyer, P. J Flynn, and S. Cheng, "The challenge of face recognition from digital point-and-shoot cameras," in *Proc.* 6^{th} *International Conference on Biometrics*, Madrid, Spain, June 4-7 2013, pp. 1–8. 3

[25] D. Baltieri, R. Vezzani, and R. Cucchiara, "SARC3D: A new 3D body model for people tracking and re-identification," in *Proc.* 16^{th} *International Conference on Image Analysis and Processing*, Ravenna, Italy, Sep. 14-16 2011, pp. 197–206. 3

[26] Yi Yang and Deva Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 20-25 2011, pp. 1385–1392. 4

[27] X. Zhou, S. Leonardos, X. Hu, and K. Daniilidis, "3D shape reconstruction from 2D landmarks: A convex formulation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, June 7-12 2015, pp. 4447–4455. 4