Face Image Quality Assessment Based on Learning to Rank

Article *in* Signal Processing Letters, IEEE · January 2015

DOI: 10.1109/LSP.2014.2347419

Face Image Quality Assessment Based on Learning to Rank

Jiansheng Chen, *Member, IEEE*, Yu Deng, Gaocheng Bai, and Guangda Su

***Abstract*—Face image quality is an important factor affecting the accuracy of automatic face recognition. It is usually possible for practical recognition systems to capture multiple face images from each subject. Selecting face images with high quality for recognition is a promising stratagem for improving the system performance. We propose a learning to rank based framework for assessing the face image quality. The proposed method is simple and can adapt to different recognition methods. Experimental result demonstrates its effectiveness in improving the robustness of face detection and recognition.**

***Index Terms*—Face quality, face recognition, learning to rank.**

I. Introduction

HUMAN face is believed to be an ideal biometric feature for personal authentication because it is universal, discriminative, non-intrusive, and easy to obtain. During the past two decades, automatic face recognition technology has attracted a great amount of attention from both academia and industry. Many methods have been proposed for identifying or verifying personal identities based on face images [1]–[3]. However, the effectiveness of automatic face recognition is challenged by variations in illumination, pose, occlusion and expression in the captured face images [4] largely because the face image acquisition process is non-contact in nature. Such problems become even more serious in real applications with uncooperative users and uncontrolled environments. Although many approaches have been proposed for improving the robustness of face recognition against different kinds of face image quality degradation [5]–[7], it is still widely understood that most face recognition methods achieve better performance on high quality face images [8]. Take face verification vendor tests for example. In the Multiple Biometrics Evaluation (MBE) organized by NIST in year 2010, on a face database consists of high quality visa photo images, the lowest error rate reported was 0.3% (False Rejection Rate at False Acceptance Rate = 0.001) [9]. However, on the LFW database [10] made up of wild face images collected from the web, the latest reported result indicates a corresponding error rate of no less than 18% [11], which is nearly two orders of magnitude worse than that in MBE.

In many practical video based face recognition systems, it is actually possible to acquire multiple face images from the target subjects. Selecting high quality face images for recognition can not only improve the system robustness and suppress false alarms, but also reduce the overall computation load considering that face feature extraction is usually complex. Berrani and Garcia were among the first to study this problem and proposed to use robust PCA for removing low quality face images as outliers [12]. This method, however, cannot be applied in situations like video surveillance in which low quality face images dominate. A more straightforward approach to solve this problem is face image quality assessment, of which most existing methods are based on the analysis of specific facial properties. Yang *et al.* adopted a tree structure for pose estimation and used the results for evaluating face image quality [13]. Gao *et al.* proposed to use the degree of facial asymmetry to quantify the face quality degradation caused by non-frontal illumination and poses [14]. Sellahewa *et al.* directly used the universal image quality index [15][16] for measuring the face image quality in terms of illumination distortion in comparison to a specified reference face image [17]. Wong *et al.* proposed a patch based probabilistic model for quality assessment trained on reference face images with frontal poses, uniform illumination and neutral expressions [18]. However, the effectiveness of these methods are limited by the applicability of the artificially defined facial properties and empirically selected reference face images. To solve this problem, we propose a simple and flexible framework for face image quality assessment, in which multiple feature fusion and learning to rank are used.

The rest of this letter is organized as follows. Section II elaborates the preprocessing of the face image. Section III presents the proposed face quality assessment method. Section IV demonstrates the application of our proposal through experiments. Section V concludes our work.

II. Face Normalization

Ideally, only image pixels inside the human face should be used for assessing face quality. This can be realized, for example, by locating contour landmarks and generating a specific

Manuscript received June 26, 2014; accepted August 06, 2014. Date of publication August 13, 2014; date of current version August 22, 2014. This work was supported by the National Natural Science Foundation of China (61101152), the Tsinghua University Initiative Scientific Research Program (20131089382), and the Beijing Higher Education Young Elite Teacher Project (YETP0104). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaokang Yang.

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: jschenthu@tsinghua.edu.cn; dengyuee@tsinghua.edu.cn; baigaowahaha@126.com; susu@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2347419

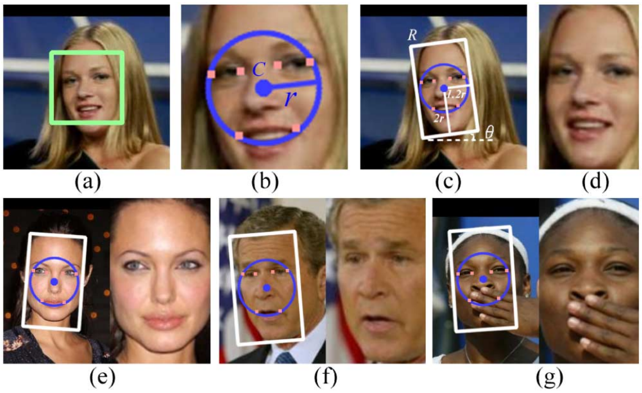

Fig. 1. Face normalization using smallest enclosing circle.

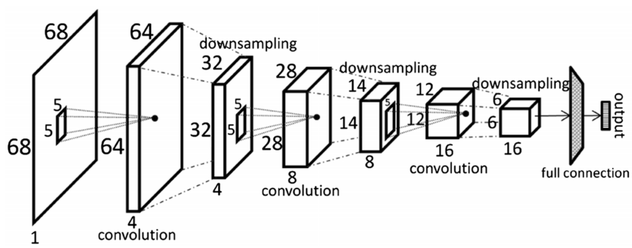

Fig. 2. Structure of the CNN for landmark location.

mask for each face in the image. However, this can be time consuming and may cause difficulties in subsequent feature extraction due to shape irregularity. On the other hand, most face detectors [19] simply output square bounding boxes which obviously deviate from human face shapes and may contain a considerable amount of non-face information. In addition, in-plane rotation of faces should not be treated as quality degradation since most face recognition systems are able to handle it properly [4]. Based on all these considerations, we propose the face normalization process shown in Fig. 1.

Fig. 1(a) shows the face detection [19] result on an image from the LFW database. The detected face area is then resized to 68×68 pixels and passed to a CNN (Convolutional Neural Network) for landmark location [20]. We use the eye/mouth corners because these landmarks are clearly defined and cover most face regions. Fig. 2 shows the structure of the CNN which contains three convolution layers and three downsampling layers. Altogether 164 convolution kernels of size 5×5 are used and the output of the network is the vector form of the landmark coordinates. We randomly select 10000 images from LFW for training and the remaining images are used for testing. The average landmark location error is 1.4 pixels on the test set. Fig. 1(b) shows the located landmarks.

To normalize the face area and eliminate in-plane rotation, we first calculate the center C and radius r of the smallest circle that enclose all the landmarks using the linear time algorithm proposed in [21]. We then place a rectangle R of size $2.4r \times 4r$ centered at C as is shown in Fig. 1(c). Obviously, all the landmarks are guaranteed to be enclosed by R. Suppose the coordinates of the four eye corners are $[x_i, y_i], (i = 1, 2, 3, 4)$, the orientation of the rectangle can be determined by equation (1), in which $\overline{x}$ and $\overline{y}$ are the mean values of the horizontal and vertical coordinates respectively. Thus, the shorter side of R is parallel to the line that best fits the four eye corners. The normalized face area in Fig. 1(d) can thus be achieved by rotating the rectangular area inside R around C by angle θ.

$$\theta = \arctan\left(\left(\sum_{i=1}^{4} x_i y_i - 4\overline{xy}\right)\right) / \left(\sum_{i=1}^{4} {x_i}^2 - 4\overline{x}^2\right)\Bigg) \tag{1}$$

More face normalization results on LFW images are shown in Fig. 1(e), (f), (g). It can be observed that the normalized faces are compact and guaranteed to contain main facial parts. The normalized faces are then used as inputs to the face quality assessment process to be introduced in the next section. The proposed normalization method is somewhat robust to inaccuracy in landmark location. Nevertheless, in case that multiple landmarks are significantly incorrectly located simultaneously, the normalization result may deteriorate and lead to a low face quality assessing result. This problem, however, can be tolerated in our work considering that such a situation, for most cases, does indicate very low face quality.

III. Face Quality Assessment

It is in general difficult to explicitly define and quantify the quality of a face image. There have been mainly two approaches for solving this problem in previous research. The first one is to empirically use certain facial properties, such as the resolution, pose angle, or illumination parameters, to quantify face image quality [13][14]. The second one is to compare a face image to selected ‘standard’ faces and use their discrepancies for measuring face quality [17][18]. Both approaches are inflexible and lack of applicability since neither of them has taken into account the possible differences among face recognition methods. For a face recognition algorithm good at solving the occlusion problem [7], Fig. 1(g) is probably more preferable than Fig 1(f). On the contrary, for a recognition method in which poses can be properly handled [6], Fig. 1(f) should be considered of higher quality. Also, face image quality should be considered in a relative manner. For most recognition methods, Fig. 1(d) is better than Fig. 1(f) but worse than Fig 1(e) in terms of face quality.

Based on the above considerations, we propose a simple and flexible face quality assessment approach based on learning to rank [22]. Suppose a face recognition method is tested on two different face databases A and B; and the recognition performance on A is better than it is on B. This indicates that for this specific recognition method, face images in A have higher quality than those in B. We note this as $A \succ B$. Let I_i and I_j be two images selected from A and B respectively; and let $f()$ be the function that transform a face image to a feature vector. Define a linear form quality assessment function $S(I) = w^T f(I)$, and our goal is to find the value of rank weight w that satisfies as many constraints in equation (2) as possible. Also, images in the same face database should be considered of similar face quality. This can be expressed by the equality constraints in equations (3) and (4). Considering the ranking nature of this formulation, we name the value of $S(I)$ as the RQS (Rank based Quality Score) of I.

$$w^T f(I_i) > w^T f(I_j); \quad \forall I_i \in A, \forall I_j \in B \tag{2}$$

$$w^T f(I_i) = w^T f(I_j); \quad \forall I_i \in A, \forall I_j \in A \tag{3}$$
$$w^T f(I_i) = w^T f(I_j); \quad \forall I_i \in B, \forall I_j \in B \tag{4}$$

The above problem formulation is identical to that in [23] and thus can be transformed into a convex max-margin formulation shown in equation (5) by introducing non-negative slack variables. λ_1, λ_2 and λ_3 are constants balancing the degree of slackness allowed by the corresponding constraints. The primal optimization problem defined by equation (5) can be efficiently solved using Newton's method [24].

$$\begin{aligned} &minmize(\|w^T\|_2^2 + \lambda_1 \sum \xi_{ij}^2 + \lambda_2 \sum \eta_{ij}^2 + \lambda_3 \sum \gamma_{ij}^2) \\ &s.t. \quad w^T(f(I_i) - f(I_j)) \geq 1 - \xi_{ij}; \quad \forall I_i \in A, \forall I_j \in B \\ &\qquad |w^T(f(I_i) - f(I_j))| \leq \eta_{ij}, \quad \forall I_i \in A, \forall I_j \in A \\ &\qquad |w^T(f(I_i) - f(I_j))| \leq \gamma_{ij}, \quad \forall I_i \in B, \forall I_j \in B \\ &\qquad\qquad \xi_{ij} \geq 0, \quad \eta_{ij} \geq 0, \quad \gamma_{ij} \geq 0 \end{aligned} \tag{5}$$

The proposed formulation can be extended to multiple databases and features. For multiple feature fusion, we use a two level learning stratagem. Suppose m different features are extracted from image I; and let the quality assessment function corresponding to the i_{th} feature be $S_i(I) = w_i^T f_i(I), (i = 1, 2, \ldots, m)$. In the 1st level learning, all the rank weights w_i are trained by solving equation (5) for different feature functions respectively. Let $\overrightarrow{S_v} = [S_1(I)S_2(I)\ldots S_m(I)]^T$ be the column vector consists of the level 1 RQS of I for different features. Define the level 2 quality assessment function of I as $S_k(I) = w_k f_\Phi(\overrightarrow{S_v})$, in which $f_\Phi()$ is the mapping function of a polynomial kernel. In our implementation, we use $m = 5$ different features and a second order polynomial kernel. As such, $f_\Phi()$ can be expressed by equation (6).

$$\begin{aligned} f_\Phi(\overrightarrow{S_v}) = [&1 \sqrt{2}S_1 S_1^2 \sqrt{2}S_2 \sqrt{2}S_1S_2 S_2^2 \sqrt{2}S_3 \sqrt{2}S_1S_3 \\ &\sqrt{2}S_2S_3 S_3^2 \sqrt{2}S_4 \sqrt{2}S_1S_4 \sqrt{2}S_2S_4 \sqrt{2}S_3S_4 S_4^2 \\ &\sqrt{2}S_5 \sqrt{2}S_1S_5 \sqrt{2}S_2S_5 \sqrt{2}S_3S_5 \sqrt{2}S_4S_5 S_5^2]^T \end{aligned} \tag{6}$$

The value of w_k can be calculated through a 2nd level training by using $f_\Phi()$ as the feature function. The value of $S_k(I)$ is then normalized to $0 \sim 100$ and is used as the final RQS of I. We use the explicit mapping function here instead of the kernel SVM to simplify RQS calculation. Fig. 3 shows the example process for calculating the RQS of Fig. 1(d).

IV. Experiments

As stated above, we suggest selecting different training images for different face recognition methods. However, to demonstrate the overall effectiveness of the proposed method, we use a heuristic criteria for data selection in our experiments. Three sets of data, $DB1$, $DB2$ and $DB3$ are prepared. $DB1$ consists of face images selected from face databases collected in controlled environments, such as FERET [26], FRGC [25] and a Chinese ID card photo database in our laboratory. $DB2$ consists of face images selected from two real world face databases: LFW [10] and AFLW [27]. $DB3$ consists of non-face natural images in which the face detector [19] generates false positive detection results. Each dataset contains 6000 images and among which 5000 are used for training and 1000 are used for testing. Fig. 4 shows examples of normalized 'face' images from the three datasets. Empirically, most face recognition methods perform better on face images collected under controlled environment than on wild face images. Therefore, in terms of face image quality, we have $DB1 \succ DB2 \succ DB3$. Five types of image features are used: HoG [28], Gabor [5], Gist [29], LBP [30] and CNN (input to the full connection layer in Fig. 2).

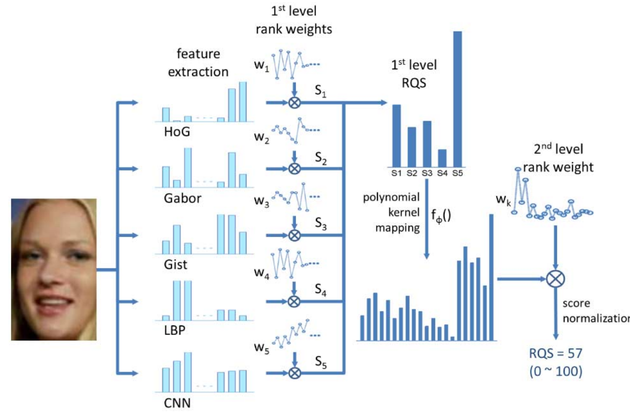

Fig. 3. Face quality assessment process.

Fig. 4. Sample face images from three data sets.

Fig. 5 shows the probability distributions of RQS value of the testing images from three datasets respectively. RQS values for sample test images are also illustrated. From left to right, the RQS value increases. It can be observed that RQS is a fairly reasonable measurement of face image quality. For example, although both of the two $DB3$ sample images are not human faces, the tiger head has a obviously higher RQS value. This is probably because that comparing to the building roof, a tiger head shares more similarity with a human face in term of both shape and structure. Also, Fig. 1(e) from LFW is very well posed and illuminated and is therefore assessed as of a pretty high quality. Its RQS value, however, is still much lower than that of the two $DB1$ images probably due to the occlusion of the forehead caused by the hair and the subtle face expression on it.

More face quality assessment results on images outside the three datasets are shown in Fig. 6, in which the RQS scores are shown at the bottom left corners of the corresponding face bounding boxes. The face of the highest RQS value in an image is indicated using a solid line bounding box. Fig. 6(a) shows the effectiveness of RQS in assessing face expressions. It can be observed that RQS reflects the intensity of expression reasonably. The highest RQS is assigned to the face image with neutral expression in Fig. 6(a). Fig. 6(b) shows two synthesized faces generated by psychologists through combining beautiful facial features from different face images [31]. The two images are well posed/illuminated and are of neutral expressions and

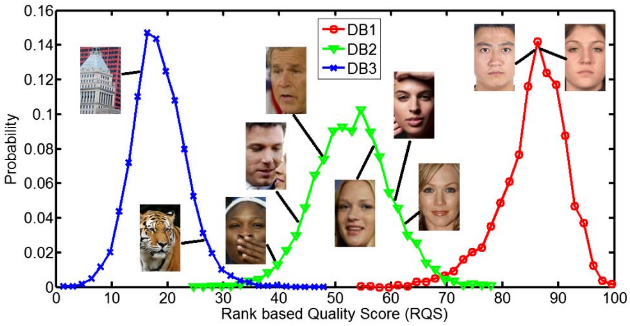

Fig. 5. RQS distributions and examples.

Fig. 6. Face quality assessment examples using RQS. The source code reproducing the experimental results shown in Fig. 6 is available at http://jschenthu.weebly.com/uploads/2/4/1/1/24110356/facequality.zip.

perfect proportions. The particularly high RQS of these two images comply well with the psychological experimental result that most people think the two faces are highly beautiful. Fig. 6(c)(d) shows the face quality assessment results on wild images in which faces vary a lot in pose, illumination, express and resolution. Fig. 6(d) demonstrates another possible application of the proposed method. For most face detectors, the number of false positive may increase when low resolution faces are to be detected. By incorporating RQS, most false detections can be effectively suppressed so as to improve the robustness of face detection.

To further verify the proposed method, we perform a face identification experiment on the SCFace database [32] which consists of real life face images of 130 subjects captured by different surveillance cameras. The face image quality varies a lot

Fig. 7. Probe image ordering according to RQS.

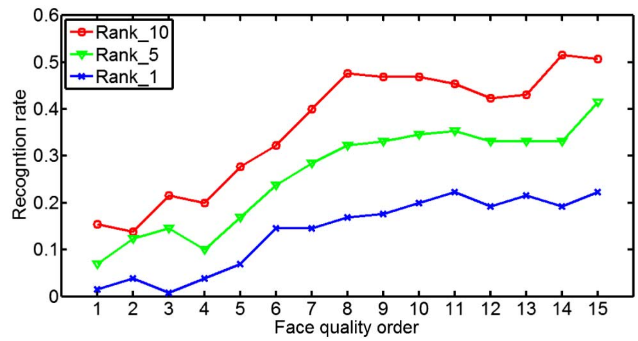

Fig. 8. Identification accuracy w.r.t. face quality ordering.

in this database due to variations in pose, expression and resolution. The database is challenging and the benchmark Rank_1 [33] recognition rate ranges from 0.7% to 7.7% [32]. We adopt the DayTime experiment setting defined in [32] in which the mug shot frontal images are used as gallery and the visible light daytime images are used as probe. Hence, for each subject, there are 15 probe images and 1 gallery image. Each probe image is compared to all the 130 gallery images for identity recognition using the Gabor filter based method proposed in [34]. For each subject, we order the 15 probe images according to their RQS values. Examples of sorted probe images are shown in Fig. 7 in which the RQS value increases from left to right. The identification accuracy is calculated with respect to the face quality ordering and the result is shown in Fig. 8. As an example, the recognition rate for the lowest quality probe image of each subject is shown at the point with the horizontal coordinate equals to 1. In general, the recognition rate increases when the face image quality improves. The Rank_1 accuracy is 22.3% for the highest quality probe images and 1.5% for lowest quality probe images. The proposed method provides an effective way for high quality face image selection in practical systems. The computational load for calculating RQS is low. A Matlab based implementation uses around 200 ms for each face image on a Quad Core desktop PC. The most time consuming part is the image feature extraction.

V. CONCLUSIONS

We propose to formulate the face image quality assessing problem in a relative manner and use learning to rank for solving it. To apply the proposed method to a specific face recognition system, we suggest selecting the training datasets accordingly for achieving better performance. For practical systems, the proposed RQS value can be used for improving face detection robustness, controlling the face quality in registration and selecting high quality images for recognition. It is also possible to use RQS for evaluating the confidence of different face images in multi-instance face recognition.

REFERENCES

[1] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[2] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 2005.

[3] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfield, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[5] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[6] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[7] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[8] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 531–543, Apr. 2007.

[9] P. J. Grother, G. W. Quinn, and P. J. Pillips, "Report on the evaluation of 2D still-image face recognition algorithms," NIST Interagency Report 7709, 2010.

[10] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. of Mass., Tech. Rep. 07-49, 2007.

[11] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," *arXiv:1403.2802*, 2014.

[12] S. A. Berrani and C. Garcia, "Enhancing face recognition from video sequences using robust statistics," in *Proc. IEEE Int. Conf. Video Signal Based Surveill.*, 2005, pp. 324–329.

[13] Z. Yang, H. Ai, B. Wu, S. Lao, and L. Cai, "Face pose estimation, and its application in video shot selection," in *Proc. Int. Conf. Pattern Recog.*, 2004, pp. 322–325.

[14] X. Gao, S. Z. Li, R. Liu, and P. Zhang, "Standardization of face image sample quality," in *Proc. Int. Conf. Biometrics*, 2007, p. 242251.

[15] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] H. Sellahewa and S. A. Jassim, "Image-quality-based adaptive face recognition," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 805–813, Apr. 2010.

[18] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection, and improved video-based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Worshop*, 2011, pp. 74–81.

[19] P. Viola and M. J. Jones, "Robust real-time face detection," in *Proc. Int. J. Comput. Vis.*, 2004, vol. 57, no. 2, pp. 137–154.

[20] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recog.*, 2013, pp. 3476–3483.

[21] E. Welzl, "Smallest enclosing disks (balls and ellipsoids)," *Lecture Notes in Computer Science 555*, pp. 359–370, 1991.

[22] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. ACM Conf. Knowledge Discovery and Data Mining*, 2002.

[23] D. Parikh and K. Grauman, "Relative attributes," in *Proc. Int. Conf. Comput. Vis.*, 2011.

[24] O. Chapelle, "Training a support vector machine in the primal," *Neural Comput.*, vol. 19, no. 5, pp. 1155–1178, 2007.

[25] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 947–954.

[26] A. P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[27] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Proc. IEEE Int. Workshop Benchmarking Facial Image Anal. Technol.*, 2011, pp. 2144–2151.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, , 2005, pp. 886–893.

[29] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[30] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recogn.*, vol. 29, pp. 51–59, 1996.

[31] C. Braun, M. Gruendl, C. Marberger, and C. Scherber, Beautycheck - Ursachen und Folgen von Attraktivitaet, [Online]. Available: http://www.beautycheck.de/cmsms/

[32] M. Grgic, K. Delac, and S. Grgic, "SCfacesurveillance cameras face database," *Multimedia Tools Appl. J.*, vol. 51, no. 3, p. 863879, 2011.

[33] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. New York, NY, USA: Springer, 2011.

[34] Y. Xiang and G. Su, "Multi-parts and multi-feature fusion in face verification," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recog. Worshop*, 2008, pp. 1–6.