AGE ESTIMATION UNDER CHANGES IN IMAGE QUALITY: AN EXPERIMENTAL STUDY

Fares Alnajar, Theo Gevers, and Sezer Karaoglu

ISLA Lab, Informatics Institute, University of Amsterdam

ABSTRACT

In this paper, we investigate the influence of image quality on the performance of aging features. Age estimation systems used or designed a number of aging features to capture the aging cues from the face such as skin texture and wrinkles. These aging cues are sensitive to small changes in the imaging conditions which suggests considering the imaging quality when extracting such information.

Although interesting performances are reported on various datasets, the effect of image quality has not been addressed. We introduce a scheme to explore the influence of image quality on the performance of appearance aging features. A number of datasets are experimented on where artifacts resulted from different types of noise are considered. Finally, we propose a method to automatically apply the most suitable features based on the quality of the image. The results show that better or comparable performance is obtained when automatically applying different features, based on image quality, in comparison to a single (best) feature type.

1. INTRODUCTION

Automatic age estimation is an important task in image processing as it has numerous applications in everyday life: i.e. security, surveillance, and online marketing. The human face reveals rich information about the age of a person and hence automatic age estimation systems are typically designed to predict the age from the human face. While aging, facial features change in response to muscle contractions and other biological changes of the skin. The facial skin becomes more leathery and rough while winkles start to appear and become more pronounced for certain face areas (e.g. around the eye and mouth corners). The shape of the face contains details indicating the age of the person. From infancy to early adulthood, the size of the head grows and certain measurements of the head size correspond to age.

A number of aging features have been designed or used to capture the aging cues. Some features [1, 2, 3, 4] are primarily applied to model the skin texture changes and the fine wrinkles on the face, while others [5, 6, 7, 8, 9] are mainly used to capture the pronounced wrinkles of the face. To measure the effectiveness of these features, they are used in age prediction pipelines.

The influence of image quality on the performance has long been a challenge in face-related image processing tasks. It is particularity important to address this challenge in age estimation since the aim here is to capture the skin texture and wrinkle changes. Moreover, the face images in real-life scenarios are taken using different capturing devices and are prone to noise due to digital transmission and compression. It is this which makes it important to study the performance of aging features with varying image quality degradation.

Many research works propose or utilize features to extract the age. A thorough survey can be found in [10]. Kwon and Labo [11] first use the head size information to infer the age. They compute ratios and measurements of distances between facial points. Other methods [12, 13, 12, 14, 15] adopt other variations of shape-based models to estimate the age. Since the shape models are limited to a certain aging period, appearance-based features are primarily utilized. They capture wrinkle changes and model the skin texture. Skin texture and fine wrinkles are represented by features like LBP [1, 2], Encoding-based Features [3] and Discrete Cosine Transformation [4]. Other gradient-based features like Gabor filters [7], Sobel [5, 6] and Biologically-Inspired Features (BIF) [8, 9] are adopted to detect pronounced wrinkles. However, since these features are designed or adjusted to capture certain aging cues, changes in image quality may introduce artifacts which influence these fine details and hence the output of these features. This suggests addressing the challenge posed by changes in image quality for age estimation.

In this paper, we investigate the effect of image quality on the performance of aging features. We mainly focus on appearance-based features. A number of age estimation datasets are used in our experiments. To investigate the robustness of the features, we simulate degrading of the image quality by systematically applying different types of digital image noise and analyze the performance. Furthermore, we introduce a basic framework to automatically assign the best aging features based on the quality of the face image.

2. THE PROPOSED SCHEME

In this section, we explain the proposed scheme used to investigate the influence of image quality on aging features performance. A key point is to isolate the feature contribution from the other factors like preprocessing and learning steps.

To this end, the datasets are reduced to similar sizes. The facial points are detected using the same landmarker and the faces are registered in the same manner. This leaves the image quality as a single variable to be tested.

To analyze the effect of the quality of images on the performance, we simulate artifacts in the datasets by applying different types of noise: JPEG-compression noise, quantization noise, and scaling noise. Each noise type is applied with different qualities to further analyze the performance. For JPEG-compression, the images are compressed with different compression rates. To simulate the quantization noise, the images are quantized to different values (lower than the original 256 pixel values). Finally, the images are scaled down by different factors and then scaled up to simulate scaling noise.

Appearance aging features can be categorized into skin-texture features and pronounced-wrinkle features. However, these features differ significantly in size, number of parameters and efficiency. For certain purposes, some features are favored over others. For instance, LBP is known to be efficient and requires a small number of parameters. This may favor it over other features like BIF which requires dense convolution of multiple Gabor filters. In this work, we focus on the details captured by the features. A suitable feature extraction method for this purpose is the Encoding-based Features [3] where local features are extracted around certain positions (e.g. in a dense manner) and the features are quantized into pre-learnt patterns. This method has the advantage that the local features can vary depending on the task. If skin texture and fine wrinkles are to be captured, sampling-like local features are deployed, while gradient-based local features are used to capture more pronounced wrinkles. Consequently, this provides an adequate tool to test the performance of features for different image qualities.

3. EXPERIMENTS

In this section, the datasets used in the experiments are first explained along with the experimental setup. We then discuss the features and the conducted experiments.

3.1. Datasets and Experimental Setup

Four publicly available dataset are used: FGNET [1], FACES [16], UvA-Nemo [17], and Morph [18]. The images in these datasets are collected from personal portrait (FGNET), or captured primarily to build the age-estimation dataset. FGNET is probably the most well-known publicly available dataset. It contains 1002 images of 82 subjects. The images are taken from personal portraits in different time periods. Morph dataset contains around 200K low quality images. A subset of 1K samples are selected. FACES dataset contains still images of 171 subjects showing six expressions. The imaging conditions are very good which allows the fine details of the face to be shown. In total, there are 1026 images in the dataset. Finally, UvA-Nemo dataset contains videos of 400 subjects showing happy expressions. The first frame, which shows a neutral face, is extracted from each video. The dataset is collected primarily for age and expression research purposes with fixed lighting conditions. Figure 1 shows example images of the four datasets.

Fig. 1. Sample images from the datasets used in our evaluation: Morph (top left), FGNET (top right), FACES (down left), and UvA-Nemo (down right).

A facial landmarker is employed to detect the eye centers which are used to register the face. The face is then cropped to the size of 125×100. All faces are converted to gray scale. The datasets are further divided into three folds of similar sizes. The identities of the subjects are mutually exclusive between the folds and the age distributions in the folds are aimed to be as similar as possible.

In Encoding-based Features, a local feature vector is calculated around each pixel. Then, the features are encoded with a pre-learnt visual codebook. A PCA tree is used to build a 256-code visual dictionary. The type of details captured is then determined by the choice of the local features. We utilize two types of features. First, sampling features where the intensities of 25 points around the pixel are sampled. This aims to capture the skin texture and fine wrinkles. Second, gradient-based features. Here, the histogram of gradient directions are computed in a 8×8 area around the pixel. This is used to model the intensity and the direction of the apparent wrinkles. The gradients are computed using Gaussian derivatives with different sigma (width) values: 0.5, 0.75, 1.0, and 1.25. The face is divided into 7×5 patches and a histogram of the visual codes is calculated for each patch. The patch-based histograms are concatenated to form the aging descriptor.

Finally, an SVM classifier is employed to learn and predict the age. A linear kernel with $C = 1$ is applied in all experiments for fair comparison and to limit the influence of the learning process. Three-fold cross-validation is used to report the results. Mean Absolute Error (MAE) is reported for the evaluation of the performance quantitatively where $MAE = \frac{1}{N}\sum_{i=1}^{N}|y^i - \hat{y}^i|$. y^i is the true age for the test sample i, $\hat{y}^i$ is the predicted age for the test sample i, and N is the number of the test samples.

[1] http://www.fgnet.rsunit.com

Fig. 2. Variations of an image when JPEG-compressed with different qualities (from left to right): 100, 75, 50, 25, and 10. Where 0 is the lowest quality and 100 is the highest quality (no compression).

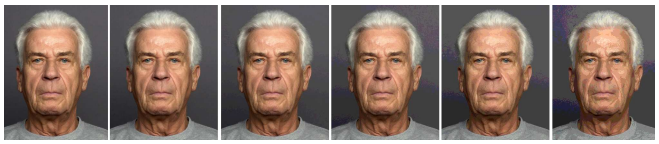

Fig. 3. Variations of an image when quantized to different values (from left to right): 256 (original) 64, 32, 16, 12, and 8.

In the following experiments, we investigate the influence of image quality degradation. More specifically, we explore the influence of three different types of image noise: JPEG compression noise, quantization noise, and image scale noise. For each of these noise types, the aging features are evaluated and analyzed.

3.2. Compression Noise

In this experiment, we explore the performance for images contaminated by compression-related artifacts. The images are compressed using the lossy JPEG compression with different qualities (0 is the lowest and 100 is the highest quality (no compression)). Figure 2 shows the variations of an image from FACES dataset when compressed with qualities 75, 50, 25 and 10.

The features are evaluated on the compressed datasets and the results are reported in Table 1. A performance is considered better than another if the error is reduced by no less than 0.1 otherwise the two performances are deemed comparable. For 75-quality and 50-quality compressions, the best results are obtained with gradient features with sigma equals to 0.5 when applied on FACES and Morph datasets. For other datasets, the performance of sampling and gradient features are comparable. When 25-quality compression is applied, gradient features with larger sigma values (0.75-1.25) give better results. This is because larger filters are required to smooth and capture the wrinkles since fine details disappear gradually. With 10-quality compression, even larger sigma values provide better gradient features than the ones with smaller sigma values or the sampling features. The results show that the more artifacts the image contains, the wider the filter is preferred to capture the aging cues.

3.3. Quantization Noise

The image intensities are typically quantized into 256 values. However, fewer quantized values can be used to reduce the image size which introduces quantization noise. In this experiment, we vary the quantized values (from 256) to 64, 32, 16, 12, and 8. Figure 3 shows the variations of an image from FACES dataset when different quantization rates are applied.

The performance of the aging features for different quantized values are shown in Table 4. The quantization noise gradually erase the aging details. Similar to applying JPEG-compression noise, the more quantization noise the image has, the wider the filter is preferred to produce better performance.

Fig. 4. Variations of an image when scaled down with different factors (from left to right): 1 (original), 2, 4, 6, and 8.

Table 1. Evaluation of aging features on FGNET, FACES, UvA-Nemo, and Morph datasets when JPEG-compressed with different qualities: 75, 50, 25, and 10 (0 is the lowest quality and 100 is the highest quality (no compression)). The number after the dataset name refers to the compression quality. e.g. FACES-75 is the dataset FACES when compressed with 75 quality. SAM refers to sampling local features and GR050, GR075, GR100, GR125 refer to the gradient-based local features for sigma values of 0.5, 0.75, 1.0, and 1.25 respectively.

Dataset	SAM	GR050	GR075	GR100	GR125
FGNET-100	7.33	7.49	7.52	7.36	7.44
FACES-100	11.05	**9.15**	10.34	10.31	11.23
UvA-Nemo-100	**6.55**	6.85	6.77	7.08	7.46
Morph-100	6.70	6.64	6.92	6.74	6.86
FGNET-75	7.39	7.50	7.48	7.49	7.39
FACES-75	11.25	**9.80**	10.21	10.76	10.82
UvA-Nemo-75	6.77	7.04	6.69	7.09	7.66
Morph-75	6.82	**6.63**	7.06	**6.67**	6.88
FGNET-50	7.56	7.48	7.48	7.55	7.54
FACES-50	10.76	**10.06**	10.21	10.97	10.99
UvA-Nemo-50	7.11	7.26	7.13	7.36	7.57
Morph-50	6.89	**6.46**	7.00	6.87	6.83
FGNET-25	7.81	**7.47**	7.65	**7.48**	7.58
FACES-25	11.98	**10.42**	11.30	11.15	11.40
UvA-Nemo-25	8.03	7.76	**7.29**	**7.36**	7.44
Morph-25	7.05	6.98	6.85	6.76	**6.57**
FGNET-10	8.89	8.31	8.06	**7.86**	**7.89**
FACES-10	13.64	12.86	**12.19**	12.32	12.32
UvA-Nemo-10	9.73	9.38	9.37	9.35	**8.83**
Morph-10	7.54	7.38	7.44	7.23	**6.98**

3.4. Scaling noise

We refer to the type of noise when scaling up an image (from a low resolution) as being scaling noise. In this experiment, we simulate scaling noise by scaling down the images by different factor and then scaling up to the original size. More specifically, the images are scaled down by 2, 4, 6, and 8 factors. Figure 4 shows an image from FACES dataset when scalled down with different factors.

Table 3 shows the performance of the aging features when scaling noise is introduced. Unlike, JPEG-compression and

Table 2. Evaluation of aging features on FGNET, FACES, UvA-Nemo, and Morph datasets when images are quantized with different values: 256(original), 64, 32, 16, 12, and 8. The number after the dataset name refers to the number of the quantized values. e.g. FACES-32 is the dataset FACES when images are quantized to 32 values.

Dataset	SAM	GR050	GR075	GR100	GR125
FGNET-256	7.33	7.49	7.52	7.36	7.44
FACES-256	11.05	**9.15**	10.34	10.31	11.23
UvA-Nemo-256	**6.55**	6.85	6.77	7.08	7.46
Morph-256	6.70	6.64	6.92	6.74	6.86
OKFGNET-64	7.39	7.40	7.56	7.51	7.46
FACES-64	11.06	**9.54**	10.47	10.79	11.14
UvA-Nemo-64	**6.58**	7.03	6.81	7.14	7.44
Morph-64	6.72	6.86	6.90	6.78	6.70
OKFGNET-32	7.71	7.65	7.62	7.58	7.52
FACES-32	11.18	**9.52**	10.35	11.00	10.97
UvA-Nemo-32	7.09	**6.68**	6.99	7.11	7.46
Morph-32	6.95	6.91	6.86	6.70	6.83
OKFGNET-16	8.43	**7.63**	7.80	7.64	7.69
FACES-16	10.73	**9.51**	10.48	10.54	10.74
UvA-Nemo-16	8.55	7.60	**7.17**	7.22	7.53
Morph-16	7.28	6.85	7.14	**6.77**	6.88
OKFGNET-12	9.14	7.97	7.80	7.63	**7.55**
FACES-12	10.69	**10.06**	10.71	10.58	10.67
UvA-Nemo-12	8.99	7.95	7.46	**7.41**	7.73
Morph-12	7.40	7.04	7.10	7.07	**6.99**
OKFGNET-8	10.55	8.46	8.73	8.44	**8.15**
FACES-8	11.96	**10.03**	11.57	11.08	11.51
UvA-Nemo-8	10.78	8.77	8.29	8.49	**8.23**
Morph-8	8.15	7.06	7.15	**7.00**	7.21

quantization noise, the aging features show similar behavior. i.e. all the aging features produce degrading performance and no clear preference can be suggested when scaling noise is introduced. This is because this type of noise influence the aging details in a similar manner and, hence, the aging features are negatively affected similarly.

3.5. Automatic Feature Assignment

In the previous section, our experiments show how the performance of the aging features depends on the quality of the images. In this experiment, we aim to automatically assign the most suitable features to each test sample based on the image quality. We restrict our experiment to JPEG-compression noise since it is the most common noise among the three tested types. More specifically, we train a classifier to detect the quality of the image and then apply the appropriate age estimator (leaned on the features most suitable for the detected quality). The results are shown in Table 4. The results here are reported for the dataset and its compressed variations (the original with the four compressed variations). We compare the obtained performance with the performance when using a single (best) feature type. For example, for FACES dataset, the single best feature type is GR050 and hence we compare the proposed method against these features for FACES dataset. Regarding UvA-Nemo dataset, GR100 features are compared against and so on. The results show that better or comparable performance is obtained with assigning the most suitable features in comparison to applying a single (best) feature type for all image qualities.

Table 3. Evaluation of aging features on FGNET, FACES, UvA-Nemo, and Morph datasets when images are scaled down with different factors: 1(original), 2, 4, 6, and 8. The number after the dataset name refers to the downscaling factor. e.g. FACES-4 is the dataset FACES when images are scaled down by factor 4.

Dataset	SAM	GR050	GR075	GR100	GR125
FGNET-1	7.33	7.49	7.52	7.36	7.44
FACES-1	11.05	**9.15**	10.34	10.31	11.23
UvA-Nemo-1	**6.55**	6.85	6.77	7.08	7.46
Morph-1	6.70	6.64	6.92	6.74	6.86
OKFGNET-2	7.41	**7.27**	7.58	7.40	7.49
FACES-2	11.87	**10.29**	10.51	11.31	11.29
UvA-Nemo-2	**6.95**	7.10	7.17	7.30	7.52
Morph-2	6.84	**6.54**	6.98	6.85	6.93
OKFGNET-4	7.55	**7.43**	7.59	7.70	7.71
FACES-4	13.38	**11.71**	11.95	12.52	12.32
UvA-Nemo-4	8.14	8.02	8.09	**7.82**	7.86
Morph-4	7.58	**7.25**	7.42	7.46	7.57
OKFGNET-6	7.88	7.92	8.04	8.20	7.93
FACES-6	14.41	**13.39**	13.35	13.38	13.39
UvA-Nemo-6	8.67	8.79	8.73	8.93	8.81
Morph-6	8.06	7.97	8.24	7.92	8.10
OKFGNET-8	**7.88**	8.16	8.34	8.00	8.29
FACES-8	14.78	13.58	**13.48**	14.11	13.64
UvA-Nemo-8	**9.18**	9.31	9.73	9.57	9.80
Morph-8	8.36	8.39	8.68	8.52	8.37

Table 4. The performance when automatically assigning the most suitable aging features based on the image quality. The results are compared against the single best feature type for each dataset. Here, each dataset includes also its compressed variations (the original dataset and the four compressed variations).

Dataset	Best Single Feature Type	Automatically Assigned Features
FGNET	7.57	7.57
FACES	10.46	10.43
Uva-Nemo	7.45	**7.31**
Morph	6.81	**6.77**

4. CONCLUSION

In this paper, the influence of image quality on aging feature performance is investigated. The motivation here is that changes in image quality affect the aging cues which primarily capture the skin texture and the wrinkles. A scheme was proposed to isolate the contribution of the features while changing the image quality. Four age estimation datasets were experimented on. The performance was studied when different types of noise artifacts were applied to the datasets.

Finally, we introduced a framework to automatically, based on the image quality, apply the most suitable feature type. The image qualities are automatically predicted. Our results show better or comparable performance when automatically applying different features, based on image quality, in comparison to a single (best) feature type.

Acknowledgment

This research was supported by the Dutch national program COMMIT.

5. REFERENCES

[1] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[2] Zhiguang Yang and Haizhou Ai, "Demographic classification with local binary patterns," in *Advances in Biometrics*, pp. 464–473. Springer, 2007.

[3] Fares Alnajar, Caifeng Shan, Theo Gevers, and Jan-Mark Geusebroek, "Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions," *Image and Vision Computing*, vol. 30, no. 12, pp. 946–953, 2012.

[4] Shuicheng Yan, Ming Liu, and Thomas S Huang, "Extracting age information from local spatially flexible patches," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 737–740.

[5] Hironori Takimoto, Yasue Mitsukura, Minoru Fukumi, and Norio Akamatsu, "Robust gender and age estimation under varying facial pose," *Electronics and Communications in Japan*, vol. 91, no. 7, pp. 32–40, 2008.

[6] Jun-Da Txia and Chung-Lin Huang, "Age estimation using aam and local facial features," in *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on*. IEEE, 2009, pp. 885–888.

[7] Feng Gao and Haizhou Ai, "Face age classification on consumer images with gabor feature and fuzzy lda method," in *Advances in biometrics*, pp. 132–141. Springer, 2009.

[8] Guowang Mu, Guodong Guo, Yun Fu, and Thomas S Huang, "Human age estimation using bio-inspired features," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 112–119.

[9] Xin Geng, Chao Yin, and Zhi-Hua Zhou, "Facial age estimation by learning from label distributions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 10, pp. 2401–2412, 2013.

[10] Yun Fu, Guodong Guo, and Thomas.S. Huang., "Age synthesis and estimation via faces: A survey," *TPAMI*, 2010.

[11] Young Ho Kwon and Niels da Vitoria Lobo, "Age classification from facial images," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. IEEE, 1994, pp. 762–767.

[12] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1178–1188, 2008.

[13] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes, "Toward automatic simulation of aging effects on face images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 442–455, 2002.

[14] Yu Zhang and Dit-Yan Yeung, "Multi-task warped gaussian process for personalized age estimation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2622–2629.

[15] Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 585–592.

[16] Natalie C. Ebner, Michaela Riediger, and Ulman Lindenberger, "Faces: database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior Research Methods*, vol. 42, no. 1, pp. 351–362, 2010.

[17] Hamdi Dibeklioğlu, Theo Gevers, Albert Ali Salah, and Roberto Valenti, "A smile can reveal your age: Enabling facial dynamics in age estimation," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 209–218.

[18] Karl Ricanek and Tamirat Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. IEEE, 2006, pp. 341–345.