

Age Classification for Pose Variant and Occluded Faces

Wei-Ta Chu
National Chung Cheng University
Taiwan
wtchu@cs.ccu.edu.tw

Wen-Long Liu
National Chung Cheng University
Taiwan
dragonliu01@gmail.com

Jen-Yu Yu
Industrial Technology Research Inst.
Taiwan
KevinYu@itri.org.tw

ABSTRACT

We extend the object class invariant (OCI) model to age classification, for pose variant and occluded faces. With the OCI model, we first localize faces from images captured in arbitrary views, and then determine the most distinctive features. Relationships between feature points and the invariant vector are described in terms of geometry and appearance information, in the form of a probabilistic model. In contrast to previous works on age classification/estimation, we emphasize that this method is especially useful for faces captured in real-world situations.

Categories and Subject Descriptors

I.4.7 [Image Processing and Computer Vision]: Feature Measurement – *feature representation, invariants*. I.4.9 [Image Processing and Computer Vision]: Applications.

General Terms

Algorithms, Performance, Human Factors.

Keywords

Object class invariant, age classification, pose variant, occlusion.

1. INTRODUCTION

Age classification for face images has been studied in years due to its potential commercial values. For example, a cigarette vendor machine should allow only adults to purchase. A shop window mounted with a screen can adaptively display commodity information according to the people standing in front of it. Although many studies have been conducted on this issue, developing reliable methods that resist to pose variation and occlusion remains a very challenging problem.

Kwon et al. [1] provides a classical method that applies exterior facial features such as eyes, nose, and wrinkle to classify faces into babies, young adults, or senior adults. Geng et al. [2] construct a representative subspace to describe aging patterns, and project an unknown face into this space to estimate its age. With the similar idea, Fu and Huang [3] further model aging patterns as manifold structures. Recently, Ni et al. [4] construct a universal multi-instance regressor to estimate human age, based on the noisy training data retrieved from the web. Rich literature survey can be found in [5].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'10, October 25–29, 2010, Firenze, Italy.
Copyright 2010 ACM 978-1-60558-933-6/10/10...\$10.00.

Although elegant methods have been proposed to model aging patterns, most studies are persecuted by illumination changes, pose variations, and occlusion on face images. Few studies have been specially conducted on faces with such variations. To detect faces under arbitrary viewpoints and occlusion, Toews and Arbel [6] propose that objects in the same class often have some invariant factors, called object class invariant (OCI), that remain consistent or similar under different viewpoints or some degree of occlusion. For example, the vector starting from the tip of the nose to the midpoint between two eyebrows may appear on faces in arbitrary viewpoints. In their work, the invariant vector is estimated based on appropriate local feature points, i.e. SIFT (scale-invariant feature transform) feature points [7], and these features are further applied to classify human gender.

In this work, we extend the framework and features proposed in [6] to conduct age classification. Faces under arbitrary viewpoints are first detected, and then the characteristics of local feature points for estimating the invariant vector are used to classify face images into predefined age groups. Relative to gender classification, finer classification is needed in age classification (in this work, five different age classes are defined). Relative to previous works on age classification/estimation, our work is much more robust to pose variant and occluded faces.

The rest of this paper is organized as follows. Section 2 describes the characteristics of OCI, how to train an OCI model, and how to use it in detecting pose-variant and occluded faces. In Section 3, visual traits based on the OCI model are extracted, and are used to conduct age classification. Section 4 reports experimental results, followed by concluding remarks in Section 5.

2. OBJECT CLASS INVARIANT MODEL

The OCI defines a vector which is invariant in the same class of objects with different viewpoints. Figure 1 shows some samples of OCIs (white arrows) on the same person's faces in different poses (top row, from [9]), and samples of human faces in different ages (bottom row, from [10]). Although faces may have significantly different appearances, the vector from the nose to the midpoint between eyebrows keeps similar in various situations. In this section, we will describe how to model OCIs and how to detect OCIs based on local feature points.



Figure 1. Samples of OCIs in faces with different poses and ages.

2.1 Probabilistic Model of OCI

In the theory proposed in [6], a face can be described as an OCI o and a set of feature points $\{m_i\}$. A feature point m_i is described by three parameters $\{m_i^b, m_i^g, m_i^a\}$. Because the way we describe feature points is more general than that in other applications, we would use *model parts* and *feature points* interchangeably in this paper. The value m_i^b is a binary variable indicating whether the feature point is present or absent. The value $m_i^g = \{x_i, \theta_i, \sigma_i\}$ denotes geometry structure of m_i , including the xy -coordinate, the orientation of main gradient, and the scale. The value m_i^a denotes feature appearance, which is described by a 128-dimensional orientation histogram (i.e. SIFT descriptor [7]). An OCI is denoted as $o = \{o^b, o^g\}$. The value o^b is a binary variable representing presence or absence of the OCI o . The value $o^g = \{x, \theta, \sigma\}$ has the parameters similar to m_i^g .

The probabilistic relationship between an OCI o and a set of features $\mathbf{m} = \{m_1, m_2, \dots, m_M\}$ can be described as

$$p(o|\mathbf{m}) = \frac{p(o)p(\mathbf{m}|o)}{p(\mathbf{m})} = p(o) \frac{\prod_{i=1}^M p(m_i|o)}{p(\mathbf{m})}, \quad (1)$$

where the second equality is derived from the assumption that features are conditional independent given the OCI. The value $p(m_i|o)$ is the likelihood that represents the relationship between individual features and the OCI. To efficiently compute the value $p(m_i|o)$, we make simplification under the following three assumptions.

- *Assumption 1 - conditional independence between feature appearance / occurrence and feature geometry given the OCI o* : This assumption is made since we consider appearance and occurrence of a feature provide no additional information about feature geometry.

$$p(m_i|o) = p(m_i^a, m_i^b|o)p(m_i^g|o). \quad (2)$$

- *Assumption 2 - conditional independence of feature appearance m_i^a and the OCI o given feature occurrence m_i^b* : This assumption is made since we consider the OCI provides no additional information with feature appearance when a feature occurs. Therefore, we can express

$$\begin{aligned} p(m_i|o) &= p(m_i^a, m_i^b|o)p(m_i^g|o) \\ &= p(m_i^a|m_i^b, o^b, o^g)p(m_i^b|o^b, o^g)p(m_i^g|o^b, o^g) \\ &= p(m_i^a|m_i^b)p(m_i^b|o^b, o^g)p(m_i^g|o^b, o^g). \end{aligned} \quad (3)$$

- *Assumption 3 - conditional independence of feature occurrence m_i^b and OCI geometry o^g given OCI occurrence o^b* : This assumption makes the previous expression as

$$\begin{aligned} p(m_i|o) &= p(m_i^a|m_i^b)p(m_i^b|o^b, o^g)p(m_i^g|o^b, o^g) \\ &= p(m_i^a|m_i^b)p(m_i^b|o^b)p(m_i^g|o^b, o^g). \end{aligned} \quad (4)$$

Under these assumptions, the probability $p(m_i|o)$ is separated into three parts: The *appearance likelihood* $p(m_i^a|m_i^b)$ describes the probability density function of model part appearance with the occurrence of a valid model part ($m_i^b=1$). Generally, it can be modeled as a multivariate Gaussian distribution model [6]. When $m_i^b=0$, it represents all invalid model parts appearance, and can be modeled as a uniform distribution.

The *occurrence probability* $p(m_i^b|o^b)$ describes the probability of occurrence of a model part with the occurrence of the OCI. Four different situations are considered, i.e. $p(m_i^b=0|o^b=0)$, $p(m_i^b=1|o^b=0)$, $p(m_i^b=0|o^b=1)$, and $p(m_i^b=1|o^b=1)$. We count the number of these situations in training data, and denote them as $\pi_i = \{\pi_i^1, \pi_i^2, \pi_i^3, \pi_i^4\}$. According to these counts, the likelihood ratio of true vs. false model parts occurrences can be expressed as

$$\frac{p(m_i^b=1|o^b=1)}{p(m_i^b=1|o^b=0)} = \frac{\pi_i^4}{\pi_i^2}. \quad (5)$$

We call this ratio *distinctiveness* of a model part, and it is an important attribute in later processes since it measures reliability of a model part [8].

The *geometry likelihood* $p(m_i^g|o^b, o^g)$ expresses the relationship between model part geometry m_i^g and OCI occurrence o^b and geometry o^g . The model part geometry m_i^g can be transformed into the OCI geometry o^g by a transformation, denoted by $t_i(m_i^g) \rightarrow o^g$. When $o^b=1$, $p(m_i^g|o^b=1, o^g)$ can be modeled as a multivariate Gaussian distribution. When $o^b=0$, o^g and $t_i(m_i^g) \rightarrow o^g$ are undefined, and we treat $p(m_i^g|o^b=0, o^g)$ as a uniform distribution.

2.2 Model Learning

Figure 2 shows the flowchart of model learning. First, we manually label the OCI for each training image, and thus geometry information o^g of each OCI can be calculated. We then extract scale-invariant features in the presentation of $\{m_i^a, m_i^g\}$. With o^g and m_i^g , the transformation information $t_i(m_i^g) \rightarrow o^g$ is estimated [6].

After feature extraction, we cluster extracted features based on their appearance and geometry information. For each extracted feature m_i , we identify two sets G_i and A_i , which consist of features that are similar to m_i in terms of geometry and appearance, respectively. The set G_i is defined as

$$G_i = \{m_j | dis_g(t_i(m_j^g), o_j^g) < T^g\}, \quad (6)$$

where $t_i(\cdot)$ is a transformation converting m_j^g to o^g . The value $t_i(m_j^g)$ denotes the OCI $o_j^{g'}$ predicted by m_j^g , based on the transformation derived from o_i^g and m_i^g . The function $dis_g(t_i(m_j^g), o_j^g)$ calculates the distance between the predicted OCI and the truth OCI. The value T^g denotes a set of threshold involving location, orientation, and scale. A simple example is illustrated in Figure 3. If $dis_g(o_j^{g'}, o_j^g) = \{\partial_x, \partial_\theta, \partial_\sigma\}$ and $(\partial_x < T_x^g \ \& \ \partial_\theta < T_\theta^g \ \& \ \partial_\sigma < T_\sigma^g)$, then m_i and m_j are geometrically similar, i.e. $m_j \in G_i$.

The set A_i consists of features that have similar appearance to m_i , and is defined as

$$A_i = \{m_j | dis_a(m_i^a, m_j^a) < T_i^{a*}\}, \quad (7)$$

where $dis_a(m_i^a, m_j^a)$ is the Euclidean distance between m_i^a and m_j^a . The threshold T_i^{a*} changes adaptively according to the ratio of number of features in G_i and A_i :

$$T_i^{a*} = \arg \max_{T_i^{a*}} \left\{ \frac{|G_i \cap A_i|}{|G_i \cap A_i|} \right\}. \quad (8)$$



Figure 2. Flowchart of model learning.

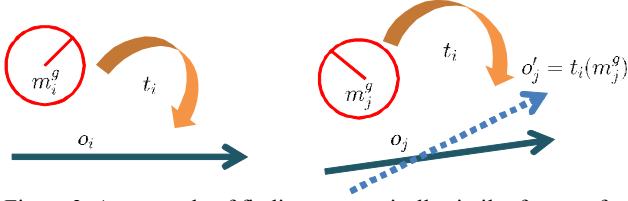


Figure 3. An example of finding geometrically similar features for the feature m_i .

The term $|G_i \cap A_i|$ denotes the number of features that are similar to m_i in terms of both geometry and appearance. The term $|\bar{G}_i \cap A_i|$ denotes the number of features that are similar to m_i in terms of appearance but not geometry. This ratio is equivalent to distinctiveness defined in eqn. (5). For each model part m_i , The features in $G_i \cap A_i$ are found.

Based on information of $G_i \cap A_i$, we remove redundant features by a method proposed in [6]. The remaining features are viewed as the representative feature set $\{m_r\}_{r=1}^N$, which are actually a subset of the set of original features in training images. Based on $\{m_r\}$, we are able to estimate the distribution of the appearance likelihood $p(m_r^a|m_r^b)$. Based on sets G_r and A_r of each model part in $\{m_r\}_{r=1}^N$, the distribution of the geometry likelihood $p(m_r^g|o_r^b, o_r^a)$ and the occurrence probability $p(m_r^b|o_r^b)$ are estimated.

2.3 Model Fitting

The goal of model fitting is to estimate the OCI in a new image by the probabilistic model described above. We start from extracting the scale-invariant features in the image, and then match the features to representative features $\{m_r\}_{r=1}^N$. If the appearance distance between an arbitrary feature m_q and a representative model part m_r is smaller than T_r^{a*} , m_q is considered to be matched with m_r . Each matched feature then infers a hypothesis o_j^g via the linear transformation learned from m_r . Therefore, many OCI hypotheses are generated from matched features. These OCI hypotheses are then clustered by a geometric clustering approach same as eqn. (6). Let's denote the result as $\{\hat{o}_1, \dots, \hat{o}_K\}$, where \hat{o}_k is the k th OCI hypothesis cluster.

Whether the k th OCI hypothesis cluster represents a true OCI instance $o^{b=1}$ or not $o^{b=0}$ is tested by a Bayes decision ratio:

$$\begin{aligned} \gamma(\hat{o}_k, \{\hat{m}_k\}) &= \frac{p(o|\{\hat{m}_k\})}{p(\bar{o}|\{\hat{m}_k\})} = \frac{p(o)}{p(\bar{o})} \prod_k \frac{p(\hat{m}_k|o)}{p(\hat{m}_k|\bar{o})} \\ &= \frac{p(o^g, o^{b=1})}{p(o^g, o^{b=0})} \prod_k \frac{p(\hat{m}_k|o^g, o^{b=1})}{p(\hat{m}_k|o^g, o^{b=0})}. \end{aligned} \quad (9)$$

The set $\{\hat{m}_k\}$ denotes the features that estimate the OCI hypotheses in \hat{o}_k . The term $\frac{p(o^g, o^{b=1})}{p(o^g, o^{b=0})}$ represents the prior ratio of the true OCI to the false OCI instance. The term $\frac{p(\hat{m}_k|o^g, o^{b=1})}{p(\hat{m}_k|o^g, o^{b=0})}$ indicates the true versus false match, which can be replaced with the distinctiveness of the model part. In our case, only one face would be in one image, and therefore we find the set of features

$\{\hat{m}_t\}$ that most probably estimate the true OCI by find the one that maximizes γ :

$$t = \arg \max_k \gamma(\hat{o}_k, \{\hat{m}_k\}). \quad (10)$$

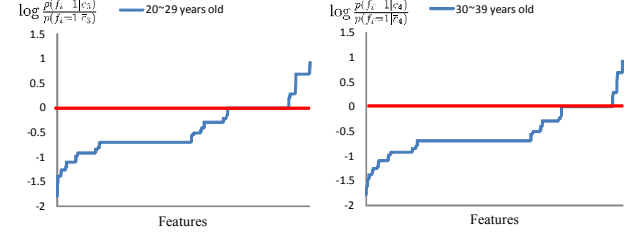


Figure 4. Log-likelihood ratio of different features in two age classes.

3. AGE CLASSIFICATION

The features $\{\hat{m}_t\}$ have been demonstrated to be effective in sex classification [6]. In this work, we exploit these features to conduct age classification. Five age classes are defined, i.e. $c = \{c_1, c_2, \dots, c_5\}$, in which $c_1 \cong 0\sim 9$ years old, $c_2 \cong 10\sim 19$ years old, $c_3 \cong 20\sim 29$ years old, $c_4 \cong 30\sim 39$ years old, $c_5 \cong 40\sim 49$ years old. To classify an image into one of the five classes based on $\{m_t\}$, a Bayesian classifier $\psi(c)$ is used, under the assumption that features in $\{m_t\}$ are conditionally independent given a class c . Thus,

$$\log \psi(c) = \log \frac{p(c)}{p(\bar{c})} + \sum_{f_i \in \{m_t\}} \log \frac{p(f_i|c)}{p(f_i|\bar{c})}. \quad (11)$$

The most possible class c^* is determined by

$$c^* = \arg \max_c \log \psi(c). \quad (12)$$

The likelihood ratio $\log \frac{p(f_i|c)}{p(f_i|\bar{c})}$ is estimated by the ratio of the number of features f_i in the training images with class c to that in images with classes other than c . In this work, we eliminate the ratio of prior probabilities $\log \frac{p(c)}{p(\bar{c})}$, because the number of training images in different classes are limited to the same number. Due to space limitation, Figure 4 only illustrates the likelihood ratios of different features in classes of 20~29 years old and 30~39 years old. Features with higher positive log likelihood ratios indicate that they are more distinctive in the corresponding class. From Figure 4, we clearly see that some features (at the rightmost) are especially distinctive, though the number of them is small relatively to other features.

4. EXPERIMENTS

We evaluate the proposed age classification method based on part of the FG-NET aging dataset [10]. There are 1002 images of 82 subjects in the ages ranging 0 to 49 years old. Faces are captured with varying illumination, pose, and expression, and may contain beards, moustaches, spectacles, or hats. The FG-NET dataset also provides 68 landmark features on each face, which is convenient for developing the method proposed in [11] that is compared with our approach. Resolution of images in this dataset is approximately 400×500 pixels. Two experiments are conducted: age classification for 1) non-occluded faces and 2) faces with an occluded circle of radii 10, 20, 40, 60 and 80 pixels. Numbers of images in different age classes and different genders are varying

in the database. Therefore, for each class we randomly select ten images for training, and test the remaining images. The experiment results are obtained from 5-fold cross validation.

4.1 Age Classification for Non-occluded Faces

We compare our approach with the method proposed in [11], which uses Active Appearance Model (AAM) to extract face appearance features and then classify features based on a Multilayer perception (MLP) neural network. Note that we experiment the method in [11] based on the truth face regions defined in the FG-NET dataset. Thus the following results for AAM+MLP are actually over optimistic. Figure 5 shows the overall F-measure of age classification for male and female images, respectively. Note that the F-measure jointly considers precision and recall rates and is defined as $F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$. In the male case, the OCI approach achieves better performances in four classes and nearly the same performance as AAM+MLP in the class of 20~29 years old. Overall, the OCI approach respectively achieves 0.26 and 0.20 F-measure for male and female images, which are superior to 0.20 and 0.19 F-measure achieved by AAM+MLP.

The performances in classes 20~29, 30~39, and 40~49 are relatively lower. The reasons for this may be: 1) few number of testing images, e.g. only nine images are tested in the class 40~49 in the female case; 2) images are misclassified into neighboring classes. For example, images in 19 years old may be misclassified into the class of 20~29. In this experiment, the FG-NET dataset doesn't have balance data in different age groups. Moreover, finer age analysis, such as age estimation, is needed for future studies.

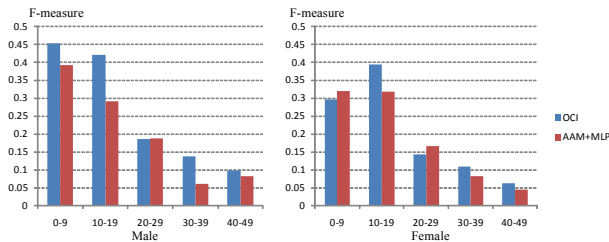


Figure 5. F-measure of age classification in (a) male and (b) female images.

4.2 Age Classification for Occluded Faces

We purposely add a black circle on the center of each test image to evaluation whether the classification methods resist to occlusion. Figure 6 shows the classification results with occluded circles with different radii. In the male case, F-measure descends when the radius of the occluded circle increases. Our OCI-based approach achieves remarkably better performance than that in AAM+MLP. In the female case, the F-measure doesn't have the clear trend as that in the male case. This may result from that the number of images in the class 10~19 is significantly larger than that in other classes, and bad classification performance in this class dominates the overall statistics. Relative to males, OCI features specific to females are less evident. This may be due to females' makeup or highly changeable hair styles. Though the performance variation in the female case is not inversely proportional to radius of the occluded circle, our OCI-based approach consistently obtains better performance than that in AAM+MLP.

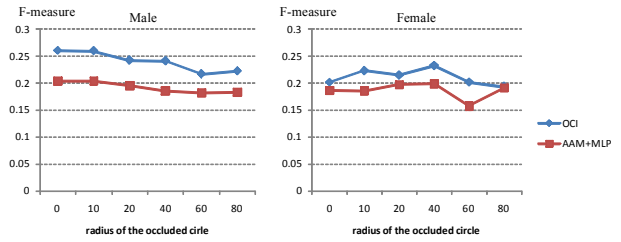


Figure 6. F-measure of age classification in occluded cases: (a) male images; (b) female images.

5. CONCLUSION

We have presented face localization and age classification for faces captured with varying illumination, pose, expression, scale, and occlusion. Such challenging work is conducted on the basis of OCI, which describes probabilistic relationship between local feature points and the invariant vector. The features that are used to estimate the OCI are demonstrated to be effective in age classification as well. We show that some features are distinctive in a specific age class, and the OCI-based approach is more reliable to localize faces and classify pose variant and occluded faces. In the future, finer analysis such as age estimation should be studied. Moreover, age estimation/classification in difference races could be studied more to facilitate more accurate results.

6. ACKNOWLEDGEMENT

This work was partially supported by the National Science Council of the Republic of China under grants NSC 98-2221-E-194-056.

7. REFERENCES

- [1] Kwon, Y.H., and da Vitoria Lobo, N. 1999. Age classification from facial images. *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1-21.
- [2] Geng, X., Zhou Z.-H., and Smith-Miles, K. 2007. Automatic age estimation based on facial aging patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234-2240.
- [3] Fu, Y., and Huang, T.S. 2008. Human age estimation with regression on discriminative aging manifold. *IEEE Trans. on Multimedia*, vol. 10, no. 4, pp. 578-584.
- [4] Ni, B., Song, Z., and Yan, S. 2009. Web image mining towards universal age estimator. *Proceedings of ACM Multimedia*, pp. 85-94.
- [5] Ramanathan, N., Chellappa, R., and Biswas, S. 2009. Computational methods for modeling facial aging: a survey. *Journal of Visual Languages and Computing*, vol. 20, pp. 131-144.
- [6] Toews, M., and Arbel, T. 2009. Detection, localization and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1567-1581.
- [7] Lowe, D.G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, vol. 60, no. 2, pp. 91-110.
- [8] Dorko, G., and Schmid, C. 2003. Selection of scale-invariant parts for object class recognition. *Proceedings of International Conference on Computer Vision*, pp. 634-640.
- [9] The color FERET database, <http://face.nist.gov/colorferet/>
- [10] The FG-NET Aging Database, <http://www.fgnet.rsunit.com/>
- [11] Wilhelm, T., Bohme, H.-J., and Gross, H.-M. 2005. Classification of face images for gender, age, facial expression, and identity. *Proceedings of International Conference on Artificial Neural Networks, LNCS 3696*, vol. I, pp. 569-574.