Fusion of Region-Based Representations for Gender Identification

Si Ying Diana Hu[1] Brendan Jou[1] Aaron Jaech[2] Marios Savvides[1]

[1]Electrical and Computer Engineering
[2]Mellon College of Science
Carnegie Mellon University
Pittsburgh, PA 15213
{sdhu,bjou,ajaech,marioss}@andrew.cmu.edu

Abstract

Much of the current work on gender identification relies on legacy datasets of heavily controlled images with minimal facial appearance variations. As studies explore the effects of adding elements of variation into the data, they have met challenges in achieving granular statistical significance due to the limited size of their datasets. In this study, we aim to create a classification framework that is robust to non-studio, uncontrolled, real-world images. We show that the fusion of separate linear classifiers trained on smart-selected local patches achieves 90% accuracy, which is a 5% improvement over a baseline linear classifier on a straightforward pixel representation. These results are reported on our own uncontrolled database of $\sim 26,700$ *images collected from the Web.*

1. Introduction

Gender classification has been a long standing challenge in the area of soft biometrics to add intelligence in security and commercial applications. A large array of classical pattern recognition and machine learning tools have been applied in this domain.

Support vector machines (SVMs) have become a particularly popular tool in the community for identifying gender [4, 14, 21, 26]. However, in past studies, evaluations of gender classification typically avoid dealing with off-angle poses, occlusions and other difficulties arising from unconstrained datasets in both training and testing. The studies typically remove pictures where the subject falls outside the range of some quality constraint (e.g. wearing glasses [8]). While generally done in good faith, these constraints heavily limit the applications of a gender classification system in real settings.

As a consequence of relying on small variations in images, studies to date have made heavy use of datasets acquired under controlled environments and that are often unfortunately small. One of the original works in the area of gender identification used a database of less than 100 images, known as SexNet [11]. More recently, the Facial Recognition Technology database (FERET), which has become the standard database for gender classification benchmarks, has just $\sim 1,010$ unique subjects [22].

As a result of the limited variation in the FERET database, the reported accuracies of other papers are relatively high in the 90%+ range [18, 12, 13, 15, 16, 17, 21, 26]. This is about 10% higher than what it would be using the same techniques on images taken from the web [5, 15]. Intuitively, a high performing classifier on a more constrained database is not guaranteed to do as well on an unconstrained database. In this study, we design a system that works well on real-world images. We develop a framework based on the fusion of multiple localized classifiers in smart-selected regions on the face. Lu et al. [19] proposed a similar region-based classifier with the purpose of addressing pose variation. Their method is composed of seven SVM classifiers trained on overlapping regions of the face. Their regions range from covering the whole face and to simply the left eye and were selected to include the "important" areas of the face at different scales to bolster performance in the presence of local variations.

In contrast, our regions are not statically determined. Instead, regions depend on the locations of key facial landmark points. We then fuse classifiers from each separate region by using probabilistic classifiers. The intent is that the framework will allow for unconstrained gender identification in the presence of pose, occlusions, illumination and scale variations. The fusion gives ability to weight certain regions of the face lower when they variations are present in those regions, and similarly, weight regions that do contribute higher. We seek to prove this concept using linear classifiers under these unconstrained circumstances, and

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

Figure 1. Sample Images from Database.

also show that performance for more constrained databases like FERET is not lost.

2. Dataset

Our dataset consists of 26,766 facial images of 13,383 females and 13,383 males which were collected from a popular image sharing website, Yahoo!®'s Flickr® [25]. The images were acquired through a series of queries designed to give a representative sample across age, gender and race. In addition, because of the nature of its source, our Flickr dataset has considerable variability in image size, illumination, pose, background scene, focal length, et cetera. As much as possible, we avoided including multiple photographs of the same subject in the database.

For each image, we collected their corresponding groundtruth gender and the location of several key landmarks for each face. The locations of the facial landmarks are presented in Figure 2 and were selected based on craniofacial anthropometry studies. We denote seven key landmarks on the face: the point on the left boundary of the left eye p^1, right boundary of the left eye p^2, left boundary of the right eye p^3, right boundary of the right eye p^4, nose p^5, center of the upper lip p^6, and chin p^7. The groundtruth labels and landmark positions were obtained using a web-based service, Amazon.com©'s Mechanical Turk (MTurk) [1]. The groundtruths and landmarks for each face were validated across three workers and we quality checked each sample manually.

Before extracting features for the classification task, we pre-process the images to normalize for the orientation and scale of each face with respect to the position of the eyes using the facial landmarks. Each face is subsequently scaled and cropped around the face to a resolution of 128×170. Several example cropped full face images of the database are shown in Figure 1.

Proper face alignment by orientation and scale normalization allows a fair comparison of gender discriminating information in the image. In general, it is not possible to

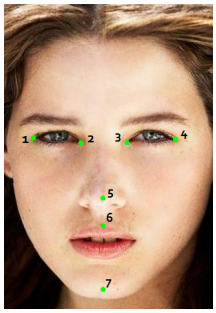

Figure 2. Selected Cranio-facial Landmarks.

align two face images such that all of the principal facial features (e.g. eyes, nose, and mouth) are also aligned with each other because each region has a different structure and shape across individuals. As a result, studies have focused on normalizing to single facial features, often opting for eye-alignment because of the low volatility with respect to position on the face. Figure 3 and Table 1 show the extent to which the relative location of each person's eyes, nose, lips, and chin vary in our Flickr database after de-rotation with respect to the eyes. We note that the point distributions shown in Figure 3 are based on the 'Full Face' crop to be discussed in Table 2 and are rotated with respect to the eyes and scaled with respect to the image to show database variations.

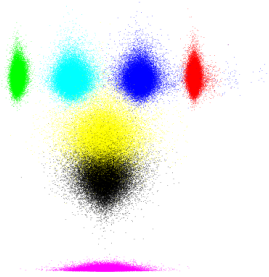

Figure 3. Point Distribution of the Landmark Points for the $26,766$ Faces in the Flickr Database.

Point	σ_X	σ_Y
p^1	1.81	5.79
p^2	4.86	6.82
p^3	5.94	6.86
p^4	1.94	5.79
p^5	10.18	11.84
p^6	7.81	9.53
p^7	7.49	1.22

Table 1. Standard Deviation in Pixels of Landmark Points from the Flickr Database at 128×170 Resolution.

The highest standard deviation observed is in point p^5 (nose) due to the changes in pose, and points p^6 and p^7 (chin and mouth, respectively) due to the variety of expressions present in images from Flickr. Additionally, some of the variation seen along the Y direction of all the points can also be attributed to changes in pitch.

3. Implementation

3.1. Region Selection & Localization

In addressing the challenges faced by variations in facial appearance, we focus on several local regions centered

Region	Left Bound	Right Bound	Top Bound	Bottom Bound
Left Eye	$p_x^1 - 0.5\left\|p_x^2 - p_x^1\right\|$	$p_x^2 + 0.5\left\|p_x^2 - p_x^1\right\|$	$\max\left(p_y^{[1-4]}\right) - \left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$	$\max\left(p_y^{[1-4]}\right) + 0.5\left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$
Right Eye	$p_x^3 - 0.5\left\|p_x^4 - p_x^3\right\|$	$p_x^4 + 0.5\left\|p_x^4 - p_x^3\right\|$	$\max\left(p_y^{[1-4]}\right) - \left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$	$\max\left(p_y^{[1-4]}\right) + 0.5\left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$
Nose	$p_x^1 + 0.5\left\|p_x^2 - p_x^1\right\|$	$p_x^4 - 0.5\left\|p_x^4 - p_x^3\right\|$	$p_y^5 - 0.9\left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$	$p_y^5 + 0.5\left\|p_y^2 - p_y^1\right\|$
Lips	p_x^1	p_x^4	$p_y^5 + 0.5\left\|p_y^6 - p_y^5\right\|$	$p_y^7 - 0.35\left\|p_y^7 - p_y^6\right\|$
Chin	p_x^1	p_x^4	$p_y^7 - 0.45\left\|p_y^7 - p_y^6\right\|$	$p_y^7 + 0.15\left\|p_y^7 - p_y^6\right\|$
Full Face	$p_x^1 - 0.5\left\|p_x^2 - p_x^1\right\|$	$p_x^4 + 0.5\left\|p_x^4 - p_x^3\right\|$	$\max\left(p_y^{[1-4]}\right) - \left\|p_y^5 - \max\left(p_y^{[1-4]}\right)\right\|$	p_y^7

Table 2. Craniofacial Struture-Dependent ROI Boundaries. The notation $p_{\{x,y\}}^{[1-7]}$ denotes x- and y-coordinates from the seven facial landmarks in Figure 2, e.g., p_x^5 references the x-coordinate of landmark $\#5$ (nose). Sub-pixel coordinates are rounded to the nearest pixel.

around the main facial features, in particular, the eyes, nose, mouth and chin. The regions may be thought of as approximately orthogonal in the sense that each region can be manipulated independently from a physiological standpoint, say, the movement of the right eye is mostly independent of the lip area. A perceptual and pyschological study by Brown et al. [3] claims that humans use the brow and eyes, brow alone, eyes alone, jaw alone, chin alone, nose and mouth, and mouth alone in order of descending importance to discriminate gender in others. And from a computational perspective, we can also expect that these regions carry higher texture information than say, the cheeks or forehead, to contribute toward gender recognition.

Our representation is motivated in part by Lu et al. [19], where they use "multi-regions" consisting of a tight face crop, the upper and lower halves of the face, left eye, nose, and mouth. Specifically, our region representation partitions the face into five regions of interest (ROIs): right eye, left eye, nose, lips, and chin.

The bounding boxes used to crop these regions as well as our full face baseline are presented in Table 2, and are function of heuristically determined ratios of different facial metrics along the face, like the lip-to-chin distance. More specifically, the boundaries of the regions use the seven face landmark points from Figure 2. In this paper, we focus on groundtruth landmarks, but there are known methods for localizing these points automatically [23]. After cropping, each ROI is then scaled to the same dimensions for comparability, as shown in Table 3.

Region	Image Size
Left/Right Eye	64×64
Nose	64×85
Lips	64×32
Chin	64×32
Full Face	128×170

Table 3. ROI Image Resolutions.

In the real-world data, the most common variations are localized occlusions or deformations, e.g. the subject has his mouth opened but the upper part of the face maintains a neutral expression, or someone may be wearing sunglasses that occludes a large region around the eyes. One of the benefits of representing the image as a collection of local regions is that we can handle these situations by relying on the correct classification in other regions that suffer less from the introducted distortion. An example of the region of interests are shown in Figure 4.

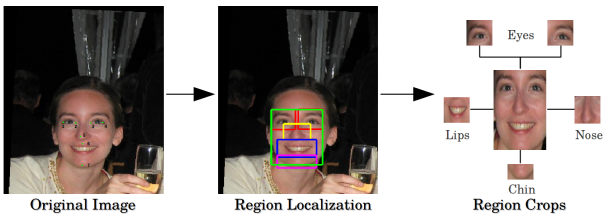

Figure 4. Regions of Interest.

3.2. Maximum Response-8 Filter Bank

The use of filter banks has long been applied in image processing to extract texture information and has found particular application to representing information about a rigid body because of the ability to design filters that are invariant to elements such as rotation and scaling. In this study, we chose to use the Maximum Response-8 (MR-8) filter set [10] in addition to baselining against raw pixel intensities. While the face cannot be assumed to be a completely rigid body because of facial variations and deformations from expressions and the like, we can still develop models that capture the regions that do satisfy these conditions.

The MR-8 filter set is based on the Root Filter Set (RFS) which consists of two anisotropic filters (one edge and one bar filter, at six orientations and three scales) and two rotationally symmetric filters (one Gaussian and one Laplacian of Gaussian, LoG). Across each rotation, the response with the highest magnitude is selected for each scale of the edge and bar filters such that eight filter responses remain in total and the set is used to characterize the original image.

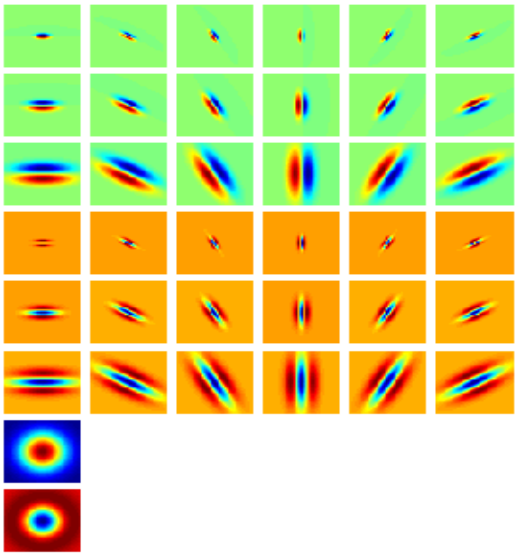

Figure 5. Root Filter Set (RFS) – Edge (green), Bar (orange), Gaussian (blue) and LoG (red) filters.

3.3. Linear Support Vector Machines

Support Vector Machine (SVM) is a powerful supervised learning algorithm used extensively in binary classification problems. SVM provides strong generalization in determining a decision boundary by finding a plane that maximizes the distance (or margin) from the closest elements (or support vectors) from the plane separating them. For N training pairs (x_i, y_i), where x_i denotes the i^{th} training sample and y_i denotes its respective class, regularization variables ξ_i, and the class mapping of $\{\text{Class A} \rightarrow y_i = 1; \text{Class B} \rightarrow y_i = -1\}$, the parameters w and C for the decision boundary $w^T X + C = 0$ can be calculated by optimizing,

$$\min_{w,C} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i.$$
$$s.t. \quad y_i(w^T x_i + C) \geq 1 - \xi_i \qquad \forall\, i \in [1, N]$$

We make use of the LIBLINEAR SVM toolbox [7] that is optimized for training on large-scale datasets.

Many recent studies on gender classification using SVMs make use of Radial Basis Functions as kernels. While the resulting accuracies are high, these classifiers are costly to train and test due to the expensive computational cost of the higher dimensional kernel projection (by Taylor series expansion, infinite even). Our proposed method seeks to prove that the use of linear classifiers is sufficient to achieve comparable accuracies. Once the descision boundaries are learned, classification is inexpensive because it only requires the dot product with weight vector and the bias, not requiring a kernel function.

3.4. Region Fusion

After the feature extraction stage, a linear SVM is individually trained for each of the five regions on the same training set. The final classification is performed through a second layer that acts as the judge to fuse the SVM margin distances from each locally trained SVM. The following techniques have been explored for this purpose:

- Majority Voting: Each region provides an equally weighted vote towards the final decision. There are no ties since there are five regions.
- Naïve Bayes: Trained on the SVM margin distances for each region, such that different weights are learned for each. The pseudo-independence of regions due to the approximately orthogonality (non-overlapping) among regions is an independence that can carry over to the conditional independence assumption.
- Logistic Regression: Trained on the distances of each region's SVM distance margins.
- Sum of SVM Distances: Decision is based on the sum of the distance margins of each SVM classifiers.

The Majority Voting is the simplest of schemes using equal weighting because we do not know in the prior which regions are more prone to variation. The Naïve Bayes and logistic regression classifiers are two different ways of fitting probabilities to the SVM response and also weighting the different regions according to their discriminating power. The last method uses a summation of SVM distance margins, giving each region an equal weight and testing if the sum is greater to less than the mean; even though SVMs are non-probabilistic classifiers, the distance to the margin of a linear SVM encodes some information about the confidence of the classification. For example, the classifier is least certain about test points that are closest to the decision boundary and most certain about points that are farther away. As a result, by the summing of the distance margins from each region, the confident regions can compensate for the regions where the classifier is incorrect and less confident.

4. Experimental Results & Discussion

Our experiments were performed on the FERET and Flickr datasets. We restricted FERET to unique subjects to avoid biases, and perform a 5-fold cross validation. For FERET, each fold was trained on 320 images from each gender and tested on 80 images per class. The accuracy with only the full face region are presented in Table 5 with a linear SVM.

The difference in accuracy rates between the two datasets highlights the variation present in real world images from the Flickr dataset not found in images taken in a

Authors	Features	Classifier	Dataset	Variation	Unique Subjects	Resolution	No. of Samples (Male/Female)	Accuracy
Proposed Method	*Region-based MR-8 filter bank*	*Fusion of Linear SVMs*	*Images from Flickr*	*Lighting, expression, pose, background, occlusion, ...*	*Yes*	*128×170*	*10037/10037 train 3346/3346 test*	*90.1% test*
			FERET [22]	*Frontal-only, studio*	*Yes*		*320/320 train 80/80 test*	*92.8% test*
D.-Y. Chen & K.-Y. Lin [5]	Color & Edge features	Adaboost + Weak classifiers	Images from Web	Lighting, expression, background, ...	Yes	Unknown	1948 total train 210/259 test	87.6% test
C.-X. Wang, Y. Liu, & Z. Y. Li [24]	Gabor filters	Polynomial-SVM	BioID	Lighting and expression	Yes	286×384	976/544 train 120 total test	92.5% test
B. A. Golomb, D. T. Lawrence, & T. J. Sejnowski [11]	Raw pixel	Neural network	SexNet [11]	Frontal-only, studio	Yes	30×30	40/40 train 5/5 test	91.9% train
S. Gutta, J. Huang, P. J. Phillips, & H. Wechsler [12]	Raw pixel	Mixture of neural networks, RBF-SVM, and decision trees	FERET	Frontal-only, studio	No	64×72	1906/1100 train 47/30 test	96.0% test
T. Jabid, Md. H. Kabir, & O. Cha [13]	Local Directional Patterns	SVM	FERET	Frontal-only, studio	No	100×100	1100/900 train ??? test	95.1% test
P.-H. Lee, J.-Y. Hung, & Y.-P. Hung [15]	Region-based	Linear regression fused with SVM	FERET	Frontal-only, studio	Unknown	Unknown	1158/615 train ??? test	98.8% test
			Images from Web	Unknown			1500/1500 train 1500/1500 test	88.1% test
X. Leng & Y. Wang [16]	Gabor filters	Fuzzy-SVM	FERET	Frontal-only, studio	Yes	256×384	160/140 total	98.1% test
			CAS-PEAL [9]	Studio	Unknown	140×120	400/400 total	93.0% test
			BUAA-IRIP	Frontal-only, studio	No	56×46	150/150 total	89.0% test
							80% train 20% test	
H. Lin, H. Lu, & L. Zhang [17]	Gabor filters	Linear SVM	FERET	Frontal-only, studio	Unknown	48×48	Unknown	92.8% test
H. Lu & H. Lin [18]	Gabor filters	Adaboost + Linear-SVM	FERET	Frontal-only, studio	Unknown	48×48	150/150 train 518 total test	90.0% test
L. Lu, X. Xu, & P. Shi [19]	Region-based	RBF-SVM fused with majority vote	CAS-PEAL	Studio	Yes	90×72	320/320 train 80/80 test	92.6% test
B. Moghaddam & M. H. Yang [21]	Raw pixel	RBF-SVM	FERET	Frontal-only, studio	Unknown	80 × 40	793/715 train 133/126 test	96.6% test
Z. Yang, M. Li, & H. Ai [26]	Texture normalization	RBF-SVM	Chinese Snapshots	Unknown	Unknown	Unknown	5600/3600 train ??? test	97.2% test
			FERET	Frontal-only, studio			1400/900 train 3529 total test	92.2% test

Table 4. Comparison of Results for Gender Identification.

Database	Feature	Accuracy
FERET	Pixel	86.6%
	MR-8	90.2%
Flickr	Pixel	78.5%
	MR-8	85.4%

Table 5. Full Face Classification Accuracy.

studio setting like in FERET. The region-specific results are shown in Table 6, where this effect is shown to be strongest in the nose region since the nose is most susceptible to appearance changes due to off-angle pose.

Region	Feature	FERET	Flickr
R. Eye	Pixel	85.8%	73.2%
	MR-8	87.2%	81.1%
L. Eye	Pixel	84.2%	69.0%
	MR-8	86.8%	81.8%
Nose	Pixel	79.9%	66.1%
	MR-8	81.4%	67.7%
Lips	Pixel	83.4%	73.2%
	MR-8	84.5%	78.9%
Chin	Pixel	73.7%	65.7%
	MR-8	78.6%	76.7%

Table 6. Region Classification Accuracy.

Note in Table 6, that although the size of each region is smaller with respect to the full face crop, the region classifiers still have significant discriminating power. Notwithstanding, no regional SVM by itself can match the accuracy of the SVM trained on the full face. As explained previously, the eye regions were selected on the assumption that the periocular region contains high gender discriminative information based on physiological studies by Brown et al. [3] and we see that this assumption is backed by the results. Also, in both the full face and region results, we see that texture-based features significantly outperform raw pixel intensities, sometimes over 10% improvements.

We fuse results from the region linear SVM distance margins, that is, before the thresholding in Table 6, to get the following results below:

Method	FERET	Flickr
No Regions	90.1%	85.4%
Majority Vote	90.9%	87.5%
Naïve Bayes	91.6%	90.1%
Logistic Regression	92.8%	89.8%
SVM Distances	91.9%	90.1%

Table 7. Fusion Classification Accuracy.

We present our results in comparison to many of the past pillars of work in gender identification in Table 4 at the top. Our use of a test set of 6692 images allows for a considerable degree of granularity of about 67 images per percent. In addition, the use of a dataset with unique subjects (i.e. without subject-biasing) and many more variations than the studio-style sets, we achieve results that much more applicable to real world inputs and also perform comparably to other previous work with studio-style datasets like FERET. With repect to our classifier, this indicates that most of the gender discriminating information is contained in local ar-

eas as opposed to the global face structure or the facial size of individual facial features. And consequently, a region-based classification framework is highly attractive because it localizes to those very interest areas.

To further support the choice of regions, we examined the decision boundary learned by Linear Discriminant Analysis and linear SVM classifiers on the full face. Linear Discriminant Analysis (LDA) is a well-known statistical method used in many high dimensional pattern recognition problems to discriminate classes. Specifically, LDA projects the data on a hyperplane that minimizes within-class scatter and maximizes between-class scatter. The projection hyperplane, often referred to as a Fisherface [2], is an important indicator of the most discriminant features of a signal from the computational perspective. As seen in Figure 6, the higher weight regions of the Fisherface corroborates our intuition that regions such as the eyes, nose, mouth, chin contain discriminating information for gender identification.

Similarly, the higher weight regions of the SVM decision boundary coincide with the discriminating portions of the Fisherface. The fact that facial hair on the chin and between the mouth and nose is a good indicator of gender and explains the high weight regions observed in those areas. Similarly, the reliance of the SVM on the area around the eyes is consistent with other studies that have shown that the periocular region is useful for discriminating gender [20].

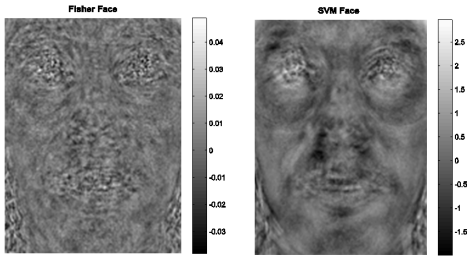

Figure 6. Fisherface weights (left) and SVM weights (right) from FERET database.

In many ways, it is intuitive why linear classifiers on the full face ROI would struggle to obtain high accuracy rates. The linear classifier would have to identify the gender discriminating information in the eyes, nose and mouth and also account for the different relative positions of these facial features and their sizes. This is a tremendous task to perform for a linear function, even in high-dimensional feature spaces. The dynamic selection of the region locations in our framework is where this a degree of non-linearity is added instead. Here, the linear SVMs trained on each region can focus on the shape and texture of each particular facial feature, capturing those local variations specifc those those regions. As observed in Figure 7, the SVM distance margins in each region provides an independent prediction of gender. And in total, the five dimensional feature space comprised of the SVM distance margins is more likely to be better separated because they are localized to their region-specific spaces.

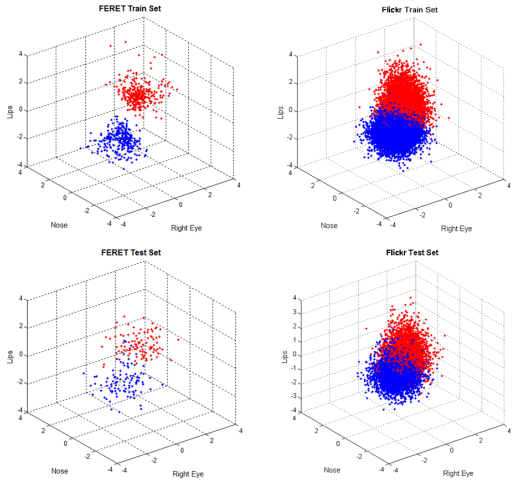

Figure 7. Vizualization of Flickr Score Distributions from the Right Eye, Nose and Lip Regions.

5. Conclusion

The proposed region-based classification method on specially selected regions of interest obtains over a 5% accuracy improvement over the full face using a linear SVM without projecting the data onto a higher dimensional plane through non-linear classifiers or kernel methods.

More significantly, our results shows that a region-based image representation handles some of the distracting variations such as pose and alignment as shown using our high-variation dataset collected from the Web. The classifier is better able to make use of the discriminating information in the image, even in the presence of common issues in real-world data like occlusions, pose, glasses and facial expressions.

Acknowledgments

We would like to thank the members of the Carnegie Mellon University (CMU) CyLab Biometrics Center for collaborating with us to collect and quality check the dataset from Flickr®. This research was funded in part by the Biometrics Identity Management Agency (BIMA) through the Army Research Lab cooperative agreement W911NF1020028.

References

[1] Amazon.com, Inc. Mechanical Turk, 2011. 2

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. volume 19, pages 711–720, 1996. 6

[3] E. Brown and D. I. Perrett. What gives a face its gender? 22(7):829–840, Feb. 1993. 3, 5

[4] L. Bui, D. Tran, X. Huang, and G. Chetty. Face gender recognition based on 2D principal component analysis and support vector machine. pages 579–582, 2010. 1

[5] D.-Y. Chen and K.-Y. Lin. Robust gender recognition for uncontrolled environment of real-life images. volume 56, pages 1586–1592, Aug. 2010. 1, 5

[6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2 edition, Nov. 2000.

[7] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, June 2008. 4

[8] H. Fukai, H. Takimoto, Y. Mitsukura, and M. Fukumi. Age and gender estimation by using facial image. In *IEEE Intl. Workshop on Advanced Motion Control*, pages 179–184, Mar. 2010. 1

[9] W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, and D. Zhao. The cas-peal large-scale chinese face database and baseline evaluations. Technical report, Joint Research & Development Laboratory for Face Recognition of the Chinese Academy of Sciences, May 2004. 5

[10] J.-M. Geusebroek, A. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. volume 12, pages 938–943, Aug. 2003. 3

[11] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. SexNet: A neural network identifies sex from human faces. In *Conf. on Advances in Neural Information Processing Systems (NIPS)*, NIPS-3, pages 572–577. Morgan Kaufmann Publishers Inc., 1990. 1, 5

[12] S. Gutta, J. R. J. Huang, P. J. Phillips, and H. Wechsler. Mixture of experts for classification of gender, ethnic orgin, and pose of human faces. volume 11, pages 948–960, July 2000. 1, 5

[13] T. Jabid, M. H. Kabir, and O. Cha. Gender classication using local directional pattern (LDP). pages 2163–2165, Aug. 2010. 1, 5

[14] A. Jain, J. Huang, and S. Fang. Gender identification using frontal facial images. pages 1082–1085, 2005. 1

[15] P.-H. Lee, J.-Y. Hung, and Y.-P. Hung. Automatic gender recognition using fusion of facial strips. page 1140, Aug. 2010. 1, 5

[16] X. Leng and Y. Wang. Improving generalization for gender classification. pages 1656–1659, Oct. 2008. 1, 5

[17] H. Lin, H. Lu, and L. Zhang. A new automatic recognition system of gender, age and ethnicity. volume 2, pages 9988–9991, June 2006. 1, 5

[18] H. Lu and H. Lin. Gender recognition using adaboosted feature. volume 2, page 646, Aug. 2007. 1, 5

[19] L. Lu, Z. Xu, and P. Shi. Gender classification of facial images based on multiple facial regions. pages 48–52, 2009. 1, 3, 5

[20] J. Merkow, B. Jou, and M. Savvides. An exploration of gender identification using only the periocular region. pages 1–5, 2010. 6

[21] B. Moghaddam and M.-H. Yang. Gender classification with support vector machines. pages 306–311, 2000. 1, 5

[22] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. of Pattern Analysis and Machine Intelligence (PAMI)*, 22:1090–1104, Oct. 2000. 1, 5

[23] K. Seshadri and M. Savvides. Robust modified active shape model for automatic facial landmark annotation of frontal faces. In *IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 319–326. IEEE Press, 2009. 3

[24] C.-x. Wang, Y. Liu, and Z.-y. Li. Algorithm research of face image gender classification based on 2-d gabor wavelet transform and svm. volume 1, pages 312–315, Dec. 2008. 5

[25] Yahoo! Inc. Flickr, 2011. 2

[26] Z. Yang, M. Li, and H. Ai. An experimental study on automatic face gender classification. In *Intl. Conf. on Pattern Recognition (ICPR)*, volume 3, pages 1099–1102, 2006. 1, 5