Eye Localization through Multiscale Sparse Dictionaries

Fei Yang, Junzhou Huang, Peng Yang and Dimitris Metaxas

Abstract— **This paper presents a new eye localization method via Multiscale Sparse Dictionaries (MSD). We built a pyramid of dictionaries that models context information at multiple scales. Eye locations are estimated at each scale by fitting the image through sparse coefficients of the dictionary. By using context information, our method is robust to various eye appearances. The method also works efficiently since it avoids sliding a search window in the image during localization. The experiments in BioID database prove the effectiveness of our method.**

I. INTRODUCTION

Accurate eye localization is a key component of many computer vision systems. Previous research disclosed that poor eye localization significantly degrades the performance of automatic face recognition systems [23]. Despite active research in the last twenty years, accurate eye localization in uncontrolled scenarios remains unsolved. The challenge comes from the fact that the shapes and appearances of eyes change dramatically under various pose and illumination. Glare and reflections on glasses, occlusions and eye blinks further increase the difficulty to this problem. Fig. 1 shows some examples from the BioID face database [18]. In these examples, the eyes are difficult to be localized.

Previous research for eye localization can be classified into three categories: geometry based approaches, appearance based approaches and context based approaches. The geometry based approaches model the eyes using geometric information. Yuille et al. [29] described eyes with a parameterized template consisting of circles and parabolic sections. By altering the parameter values, the template is deformed to find the best fit to the image. Bai et al. [3] applied radial symmetry transform to determine the eye centers. A recent work of this category is Valenti et al.'s Isophote Curvature method [27].

In the appearance based approaches, eyes are described by various photometric features including gradients [19], projections [30], edges maps [2], etc. Many statistical classification methods have been applied to model eye appearances. For instance, principal component analysis (eigeneyes) [22], support vector machines [4] [5] [17] [26], multilayer perceptrons [18], neural networks [15], and boosting methods [7] [20] [21], etc. Everingham et al. [14] compared several algorithms and found that the simple Bayesian model outperforms a regression-based method and a Boosting-based method.

This work was supported by the National Space Biomedical Research Institute through NASA NCC 9-58

Fei Yang, Junzhou Huang and Peng Yang are with the Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, 08854, USA, feiyang@cs.rutgers.edu, jzhuang@cs.rutgers.edu, peyang@cs.rutgers.edu

Dimitris Metaxas is with Faculty of the Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, 08854, USA, dnm@cs.rutgers.edu

Fig. 1. Various eye appearances (from BioID database)

The context based approaches incorporate the interaction among objects to disambiguate appearance variation. Active shape models (ASM) [10] and active appearance models (AAM) [9] localize facial landmarks (including eye landmarks) by using a global shape constrain. Cristinacce et al. [11] used pairwise reinforcement of feature responses and a final refinement by AAM. Tan et al. [24] built the enhanced pictorial structure model for eye localization.

Although there has been extensive research for eye detection and localization, reliable eye localization in uncontrolled scenarios is still far from being resolved. In uncontrolled scenarios, the geometric structures and eye appearances may be dramatically different from predefined templates or the models learned from the training data. In this case, the accuracy of most previous methods would decrease significantly. In this paper, we present a new eye localization method based on Multiscale Sparse Dictionaries (MSD). We built a pyramid of dictionaries that models context information at multiple scales. The localization algorithm starts from the largest scale. At each scale, an image patch is extracted from the previously estimated eye position. We use the sparse dictionary to reconstruct the image patch. The relative location of this patch to the eye is estimated as the position with minimum residual error. The relative location is then used to update the estimations of eye positions.

In our approach, the dictionary of each scale captures the context information of a specific range. Using large context is robust to the variation of eye appearances, and using small context enables more accurate localization. By using context information of multiple scales, our algorithm works both robust and accurately. Our method avoids sliding a search window in the image, thus is more efficient than widely used sliding window methods. Based on sparse representation and

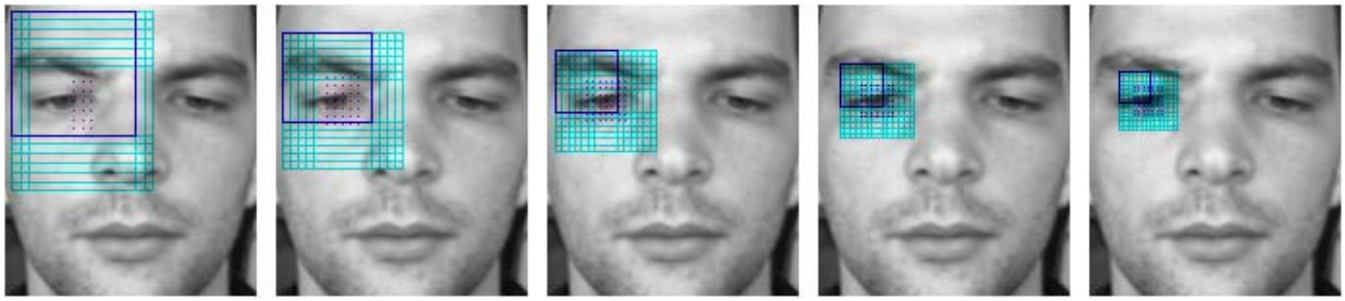

Fig. 2. Examples of multiscale dictionaries

optimization methods, the algorithm works efficiently and is resistant to image noises.

The rest of this paper is organized as follows. Section 2 introduces sparse representations, and the method we build the eye context dictionaries. Section 3 describes the localization algorithm. Section 4 shows experiment results and comparison with other methods. Section 5 concludes this paper.

II. Multiscale Sparse Dictionaries

Recently, the research of computing sparse linear representations with respect to an over complete dictionary has received increasing attention. Sparse methods have been successfully applied to a series of computer vision problems, such as image denoising [13] and face identification [28].

The sparse theory is magnetic as it implies that the signal $x \in R^n$ can be recovered from only $m = O(k \log(n/k))$ measurements [6] if x is a k-sparse signal, which means that x can be well approximated using $k \ll n$ nonzero coefficients under some linear transform. The problem can be formulated with l^0 minimization:

$$x_0 = \arg\min ||x||_0, \quad \text{while } ||y - Ax||^2 < \varepsilon \tag{1}$$

where $|| \cdot ||_0$ denotes the l^0-norm which is the number of nonzero entries and ε is the error level. Inspired by the recent work of Wright et al. [28], we first assume that the image patches at the same location do lie on a subspace. Given sufficient training patches $v_{p,i}$ at the location p, we set

$$A_p = [v_{p,1}, v_{p,2}, \cdots, v_{p,N}] \tag{2}$$

If a testing patch y is also extracted with the same size from the same location p, it should approximately lie in the linear span of the training patches associated with location p:

$$y = \alpha_{p,1} v_{p,1} + \alpha_{p,2} v_{p,2} + \cdots + \alpha_{p,N} v_{p,N} \tag{3}$$

Since the location of the test sample is initially unknown, we define a new matrix A as the concatenation of all the patches extracted from N training images at all locations:

$$A = [A_1, A_2, \cdots, A_P] \tag{4}$$

$$= [v_{1,1}, \cdots, v_{1,N}, \cdots, v_{P,1}, \cdots, v_{P,N}] \tag{5}$$

Then the linear representation of y can be rewritten in terms of all training patches,

$$y = Ax_0 \tag{6}$$

where $x_0 = [0, ..., 0, \alpha_{p,1}, \alpha_{p,2} ..., \alpha_{p,N}, 0, .., 0]^T$ is a sparse coefficient vector whose entries are zero except those associated with the location p.

The scale of the local patches is an essential factor in sparse representation. Large patches contain more context information, thus are more robust to variation of eye appearances. Small patches contain small context, and are more accurate to localize eye centers. Fig. 3 shows how we combine multiscale context information for eye localization. We start from the largest context, which gives an estimate of eye location. Subsequent dictionaries are used in smaller region and are expected to provide a closer eye location. By sequentially applying dictionaries from the largest scale to the smallest scale, the estimated eye location converges to the true position.

To build dictionaries at multiple scales. All training images are carefully aligned using the centers of eyes. The training set are further expanded by rotating and resizing while keeping the eye center fixed. We build dictionaries for left eye and right eye separately. The training patches are extracted by moving a fixed size window around the eye, as shown in Fig. 2. At each position, a patch is extracted from the ith image, and forms a column vector $v_{p,i}$. All these vectors are concatenated as A_p in Equation 2.

The dictionary A_p can be further compressed via K-SVD algorithm [1], which is an iterative method that looks for the best dictionary to represent the data samples. The training procedure is summarized in Algorithm 1.

III. Eye Localization

For a testing image, we first use a face detector to find the bounding box of the face. The face region is then cropped and normalized to be the same size of the training faces. The pixels are concatenated into a vector y, which is normalized to have unit length.

The average eye localization of the training faces $L_0 = [x_0, y_0]^T$ are used as initial estimate of the eye location. The localization procedure starts from the largest scale ($s = 1$). Orthogonal matching pursuit algorithm [25] is applied to solve the l^0-norm problem to find k nonzero coefficients.

$$x = \arg\min ||Ax - y||_2 \tag{7}$$

$$\text{subject to } ||x||_0 \leq k \tag{8}$$

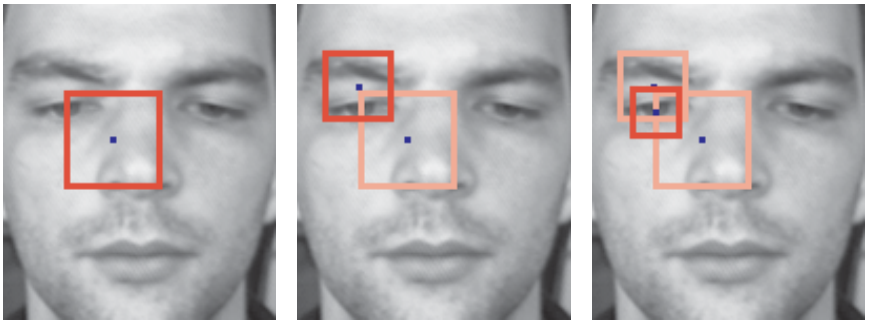

Fig. 3. Eye searching in multiple scales

Algorithm 1 *Training sparse dictionaries*

Align face images using the positions of two eyes.
Expand training set by scaling and rotation.
for scale $s = 1 : S$ **do**
  for position $p = 1 : P$ **do**
    Extract image patches at scale s and location p.
    $A_{s,p} = [v_{s,p,1}, v_{s,p,2}, \cdots, v_{s,p,N}]$.
    Normalize columns of $A_{s,p}$ to have unit length.
    Compress $A_{s,p}$ by K-SVD.
  end for
  Concatenate all the dictionaries at size s.
  $A_s = [A_{s,1}, A_{s,2}, \cdots, A_{s,P}]$
end for

The residual for each non-zero coefficient is computed as

$$r_i(y) = ||y - Ax_i||_2, \quad (i = 1, \cdots, k) \tag{9}$$

The position of current image patch is estimated as the one corresponding to the minimum residual

$$L_y = \arg \min_{\mathrm{i}} r_i(y) \tag{10}$$

For image patch y, the previous estimate for its position is L_{s-1} and the new estimate is L_y. We can update the estimate for eye position as

$$L_s = L_{s-1} + L_0 - L_y \tag{11}$$

A new image patch is then extracted at scale $s + 1$ from location L_s. The previous steps are repeated for each scale. Our localization algorithm is summarized in Algorithm 2.

As shown in Fig. 4, we have an estimated location at each scale. By applying the kernel density estimation and mean shift algorithm [8], the final estimated eye location is the position that maximize the following density function

$$f(L) = \frac{1}{S} \sum_{i=1}^{S} K\left(\frac{L - L_i}{scale_i}\right) \tag{12}$$

Algorithm 2 *Eye Localization*

Detect and crop face region.
Set initial eye position L_0.
for $s = 1 : S$ **do**
  Apply OMP algorithm to find k sparse coefficients $j_1, \cdots, j_k$
  Find minimum residual and estimate the location of current patch following Equation (9) and (10).
  Update current estimated eye position L_s following Equation (11).
end for
Repeat the above steps for the other eye.

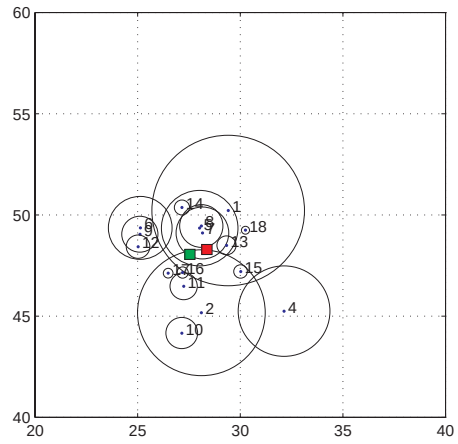

Fig. 4. Kernel density estimation to estimate eye location

IV. Experiment

In order to assess the precision of eye localization, we apply the normalized error measure introduced by Jesorsky et al. [18]. The normalized error is measured by the maximum of the distances d_l and d_r between the true eye centers C_l, C_r, normalized by the distance between the expected eye centers. This metric is independent of scale of the face and image size:

$$d_{eye} = \frac{max(d_l, d_r)}{||C_l - C_r||} \tag{13}$$

We test the precision of our algorithm in the BioID face

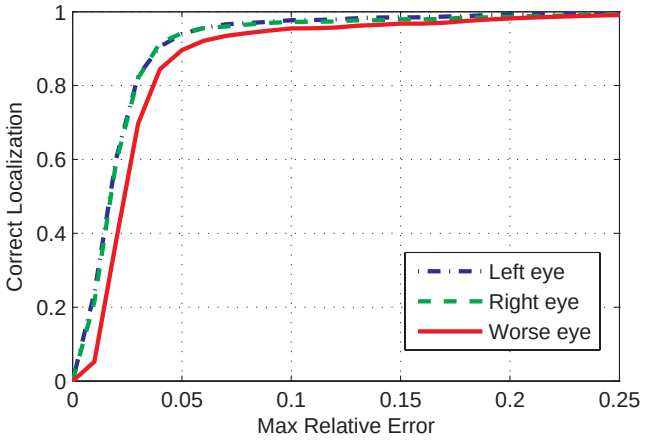

Fig. 5. ROC curve of eye detection in BioID database

BioID	$e < 0.05$	$e < 0.10$	$e < 0.25$
Jesorsky 01 [18]	40.00%	79.00%	91.80%
Hamouz 04 [16]	50.00%	66.00%	70.00%
Hamouz 05 [17]	59.00%	77.00%	93.00%
Cristinacce 04 [12]	56.00%	**96.00%**	98.00%
Asterialdis 06 [2]	74.00%	81.70%	97.40%
Bai 06 [3]	37.00%	64.00%	96.00%
Niu 06 [21]	78.00%	93.00%	95.00%
Campadelli 06 [4]	62.00%	85.20%	96.10%
Campadelli 09 [5]	80.70%	93.20%	95.30%
Valenti 08 [27]	**84.10%**	90.85%	**98.49%**
Ours	**89.60%**	**95.50%**	**99.10%**

TABLE I
COMPARISON OF EYE LOCALIZATION METHODS IN BIOID DATABASE

database [18]. The BioID database consists of 1521 frontal face images of 23 subjects. The images are taken under various lighting conditions in complex backgrounds. Thus this database is considered one of the most difficult databases for eye detection tasks.

We run two-fold cross validation and compare our results with previous methods which report the normalized errors in the same database. The methods we compare with include those used by Jesorsky et al. [18], Hamouz et al. [16] [17], Cristinacce et al. [12], Asteriadis et al. [2], Bai et al.[3], Niu et al. [21], Campadelli et al. [4] [5] and Valenti et al. [27]. For those are inexplicitly reported by the authors, the results are estimated from the graphs in paper.

Table 2 compares our results with previous methods for an allowed normalized error of 0.05, 0.1 and 0.25 respectively.Our results and previous best reported results are highlighted in bold text. Specifically, for an allowed normalized error at 0.05 and 0.25, our eye localization algorithm outperforms all previous reported results. And for an allowed normalized error at 0.10, our result is close to the best reported.

V. CONCLUSION

In this paper, we address the eye localization problem as a sparse coding problem. By assuming that an testing image patch is a linear combination of the training patches at the same position, we propose a new eye localization method by solving sparse coefficients of an over complete dictionary. In the proposed method, we build multiple dictionaries to model context of eyes at multiple scales. Eye locations are estimated from large to small scales. By using context information, our method is robust to various eye appearances. The method also works efficiently since it avoids sliding a search window in the image during localization. The experiments in BioID database prove the effectiveness of our method.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.

[2] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas. An eye detection algorithm using pixel to edge information. In *Proc. the 2nd International Symposium on Control, Communications, and Signal Processing*, 2006.

[3] L. Bai, L. Shen, and Y. Wang. A novel eye location algorithm based on radial symmetry transform. In *Proc. International Conference on Pattern Recognition (ICPR)*, pages 511–514, 2006.

[4] P. Campadelli, R. Lanzarotti, and G. Lipori. Precise eye localization through a general-to-specific model definition. In *Proc. British Machine Vision Conference*, page I:187, 2006.

[5] P. Campadelli, R. Lanzarotti, and G. Lipori. Precise eye and mouth localization. *International Journal of Pattern Recognition and Artificial Intelligencece*, 23(3):359–377, 2009.

[6] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006.

[7] L. Chen, L. Zhang, L. Zhu, M. Li, and H. Zhang. A novel facial feature localization method using probabilistic-like output. In *Proc. Asian Conference on Computer Vision (ACCV)*, 2004.

[8] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[11] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *Proc. British Machine Vision Conference*, 2006.

[12] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *Proc. British Machine Vision Conference*, 2004.

[13] M. Elad and M. Aharon. Image denoising via learned dictionaries and sparse representation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 895–900, 2006.

[14] M. Everingham and A. Zisserman. Regression and classification approaches to eye localization in face images. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 441–448, 2006.

[15] C. Garcia and M. Delakis. Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE Trans. Pattern Analyssi and Machine Intelligence*, 26(11):1408–1423, 2004.

[16] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, and H. Kälviäinen. Affine-invariant face detection and localization using gmm-based feature detector and enhanced appearance model. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 67–72, 2004.

[17] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kälviäinen, and J. Matas. Feature-based affine-invariant localization of faces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(9):1490–1495, 2005.

[18] O. Jesorsky, K. J. Kirchberg, and R. Frischholz. Robust face detection using the hausdorff distance. In *Proc. International Conference Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 90–95, 2001.

Fig. 6. Samples of localization results

[19] R. Kothari and J. Mitchell. Detection of eye locations in unconstrained visual images. In *Proc. International Conference on Image Processing (ICIP)*, pages III: 519–522, 1996.

[20] Y. Ma, X. Ding, Z. Wang, and N. Wang. Robust precise eye location under probabilistic framework. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 339–344, 2004.

[21] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao. 2d cascaded adaboost for eye localization. In *Proc. International Conference on Pattern Recognition (ICPR)*, pages 1216–1219, 2006.

[22] A. P. Pentland, B. Moghaddam, and T. E. Starner. Viewbased and modular eigenspaces for face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994.

[23] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang. Curse of misalignment in face recognition: Problem and a novel misalignment learning solution. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 314–320, 2004.

[24] X. Tan, F. Song, Z. Zhou, and S. Chen. Enhanced pictorial structures for precise eye localization under uncontrolled conditions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1621–1628, 2009.

[25] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Information Theory*, 50(10):2231–2242, 2004.

[26] M. Türkan, M. Pardàs, and A. E. Çetin. Human eye localization using edge projections. In *Proc. the 2nd International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 410–415, 2007.

[27] R. Valenti and T. Gevers. Accurate eye center location and tracking using isophote curvature. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Analysis and Machine Itelligence*, 31(2):210–227, 2009.

[29] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.

[30] Z.-H. Zhou and X. Geng. Projection functions for eye detection. *Pattern Recognition*, 37(5):1049–1056, 2004.