LEARNING SCALE RANGES FOR THE EXTRACTION OF REGIONS OF INTEREST

Qi Li, Zachary Bessinger

Western Kentucky University
Department of Mathematics and Computer Science
1906 College Heights Blvd, Bowling Green, KY

ABSTRACT

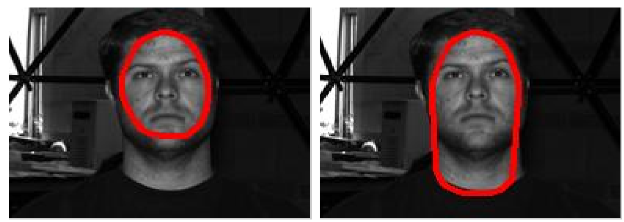
Scale space has been widely used in various applications. Given an application, it is essential to decide optimal scales under a certain criterion. Subsampling a scale space is a popular scheme to reduce the search space and thus computational costs. In the context of the extraction of Regions of Interest, we will introduce an alternative scheme that aims to learn scale ranges from training images in order to reduce the search space. We test the proposed scheme in a case study of face localization, and obtain promising results.

***Index Terms*—** Region of interest, scale, classification

1. INTRODUCTION

Scale space [6] has been shown an effective model of human vision. It has been widely used in various applications of object detection [9, 8, 11, 4, 7]. Scale selection, i.e., selection of the optimal scale, is the an important issue in the application of the theory of scale space. Scale selection can be unsupervised (without training images), e.g., in terms of local extrema [6]. Scale selection may also be supervised, where the optimality is decided by a specific tagged data. In general, it is computationally expensive to select an optimal scale in a full continuous scale space. Subsampling is thus a realistic scheme to address the computational complexity [6].

In this paper, we will introduce an alternative scheme to reduce the computational cost, which is based on learning scale ranges from a set of tagged data. Our study is performed in the context of the extraction of Regions of Interest (ROI), where we assume that an image contains a single salient attention only. Note that this assumption is valid in many applications, including: i) extraction of targeting objects from high throughput medical images, ii) initialization step of object tracking, etc. Fig. 1 shows two ROIs that are the largest ones under two different scales (10 and 13), respectively. Subjectively, the first ROI represents a face region, while the second ROI represents a region of skin (a face and a neck). Fig. 1 illustrates the motivation of introducing a learning scheme (or using tagged data) for the selection of optimal scale.



(a) (b)

Fig. 1. ROIs in two different "subjective" scales: (a) $\sigma = 10$; (b) with $\sigma = 13$.

Specifically, we will propose a learning scheme that consists of two steps. The first step is to tag scale ranges with a supervised criterion on the ROI extraction. The basic idea of supervised criteria is to maximize the consistency between a ROI and a specific labeled data (ground truth) in the context of ROI extraction. For example, for face localization where face regions are labeled, the scale $\sigma = 10$ is then considered to be better than $\sigma = 13$. The second step is on construction of feature vector space for classification. In this paper, the feature vector space is constructed via Gabor filters [2]. Furthermore, Linear Discriminant Analysis [1] is applied to extract discriminant features for scale classification.

We test the proposed scheme in a case study of face localization across poses and illuminations that can benefit the study of face recognition under unconstrained conditions [13]. We use YaleB as the test dataset that include 9 poses and 64 illumination conditions. The experimental results show the effectiveness of the proposed scheme.

The rest of the paper is organized as follows: In Section 2, we propose a learning scheme for the classification of scale ranges. Experiments are presented in Section 3. Conclusions and future work are presented in Section 4.

2. A LEARNING SCHEME

In this section, we will propose a learning scheme for the classification of scale range, as illustrated in Fig. 2. The scheme consists of two steps. The first step tags training images via

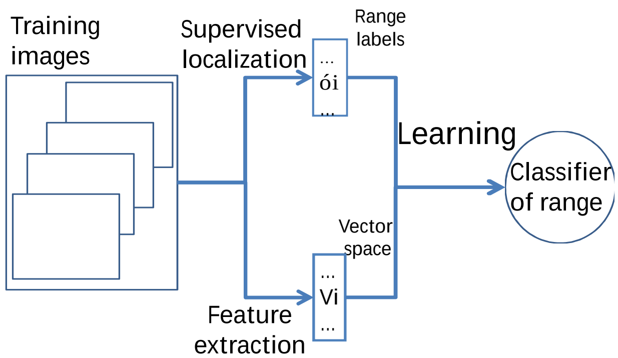


Fig. 2. A learning scheme of classification

supervised criteria. Given a training set of images with manually extracted ROI, this step tags the scale range of each image via supervised criteria on the optimization of algorithmic ROIs. The second step constructs a space of feature vectors, aiming for classification. Gabor features will be employed in this step.

For convenience, we introduce the following notations. Denote I an image, $I_\sigma = I * G_\sigma$ (where G_σ is a Gaussian filter with scale σ), R_I the ROI of ground truth, $\mathcal{R}_{I_\sigma}$ a union of all connected components of I_σ, and R_{I_σ} the largest connected component in $\mathcal{R}_{I_\sigma}$. To distinguish the ground truth ROI R_I (manually extracted from an image), we call connected components in a scale space as *algorithmic ROIs*.

2.1. Scale selection via supervised criterion

Our basic strategy of scale selection is to establish a certain optimality criterion on connected components of an image in scale space, where connected components can be obtained from an edge image. Fig. 3 shows connected components in a sampled scale space of the face image presented in Fig. 1. The scale increases from 1 to 25 with 3 as the sample step from left to right and from top to down. The largest connected component is highlighted with a red closing curve. Visually, the largest connected component associated with $\sigma = 7$ and $\sigma = 9$ are the ones most consistent with the face region, precisely speaking, the ground truth ROI R_I.

Given a scale σ, we formally define the *consistency* between algorithmic $\mathcal{R}_{I_\sigma}$ and the ground truth ROI R_I as follows:

$$consistency(\mathcal{R}_{I_\sigma}, R_I) = \max_{R \in \mathcal{R}_{I_\sigma}} \frac{R \cap R_I}{R \cup R_I}. \qquad (1)$$

We propose a criterion that maximize the consistency between $\mathcal{R}_{I_\sigma}$ and the ground truth ROI R_I, as follows:

$$\sigma^* = \text{argmax}_\sigma consistency(\mathcal{R}_{I_\sigma}, R_I). \qquad (2)$$

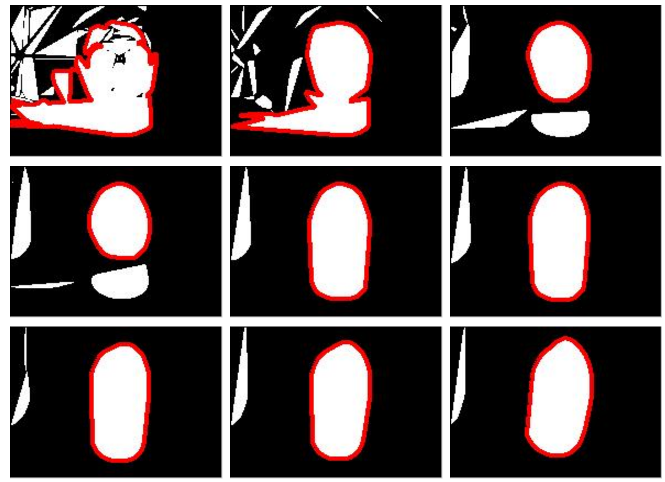


Fig. 3. An example of connected components in a (sampled) scale space, from small to large scales. The largest connected component is highlighted with a red closing curve.

Fig. 4 summarizes the supervised criterion for scale selection, In this paper, we define a full scale space as an interval $[0, \min(r, c)]$, where r and c are the dimensions of an input image.

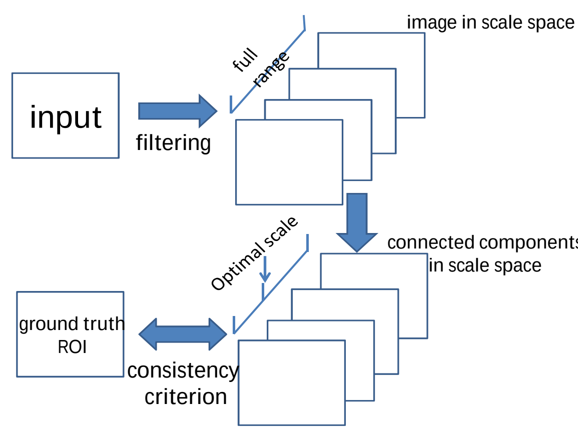


Fig. 4. Supervised criterion for scale selection by maximizing the consistency between algorithmic ROIs and the ground truth ROI.

2.2. Structuring a scale space of training images

Given a set of optimal scales $\{\sigma_i | i = 1, \ldots, m\}$ obtained from a set of training images. Denote $\sigma_{\min} = \min_i\{\sigma_i\}$, and $\sigma_{\max} = \max_i\{\sigma_i\}$. We define its scale space as the interval $[\sigma_{\min}, \sigma_{\max}]$. We propose a scheme to split the scale space into two ranges: i) *small scales*, and ii) *large scales*. Without loss of generality, we assume $\sigma_1 \le \sigma_2 \le \cdots \le \sigma_m$. Define $i^* = \text{argmax}_i |\sigma_{i+1} - \sigma_i|$. We define the *small-scale range* as

$[\sigma_{\min}, \sigma_{i*}]$, and the *large-scale range* as $[\sigma_{i^*+1}, \sigma_{\max}]$.

For example, assume that we have a scale set $\{7, 10, 22, 25\}$. The scale space is the interval $[7, 25]$. σ_{i^*} is 10 since 22-10=12 is the maximum consecutive subtraction. Therefore, the scale space is split into small-scale range [7,10], and large-scale range [22, 25].

The advantage of the above structuring method lies in maximization of the distinction of two ranges of scales, which in turn increases the separability of features during the classification. A scale range can be recursively split using the same strategy, leading a multi-class classification problem. But, we will be focused in binary classification in this paper for the convenience of illustration.

2.3. A vector space of features for classification

We construct a vector space of features from images. We will use Gabor filters to construct features.

Denote G_{σ_i,θ_i} a Gabor filter of a specific scale σ and orientation θ_i, where $i = 1, \dots, n$. Denote I_{σ_i,θ_i} the convolution of image I and the Gabor filter, and $\|I_{\sigma_i,\theta_i}\|_2$ the 2-norm of the convolved image, which represents an overall response to a Gabor filter of scale and orientation. By collecting responses to various Gabor scales and orientations, we have a vectorized representation of an image as follows:

$$(\|I_{\sigma_1,\theta_1}\|_2, \|I_{\sigma_2,\theta_2}\|_2, \dots, \|I_{\sigma_n,\theta_n}\|_2). \tag{3}$$

We apply Linear Discriminant Analysis (LDA) to extract the most discriminant feature. Note that what we are targeting is a binary classification problem, and thus the dimension of the most discriminant feature is just 1. Specifically, we extract the most discriminant features by computing an optimal discriminant projection according to the Fisher criterion:

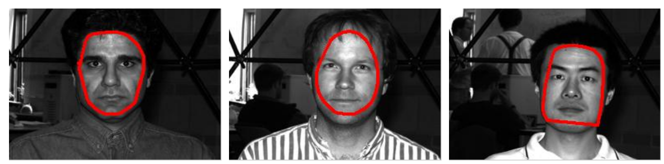
$$W^* = \text{argmax}_W \frac{W^T S_b W}{W^T S_w W}, \tag{4}$$

where S_b and S_w are between-class scatter and within-class scatter matrices, respectively. To build a classifier, we model discriminant features of small and large scales as a Gaussian Mixture Model (GMM), specifically, a mixture of two Gaussians (G_1 and G_2). We apply EM algorithm [3] to compute the parameters of GMM. An image is classified as a small scale if $G_1(v_p) > G_2(v_I)$, where v_I is the discriminant feature of I.

3. EXPERIMENTS

In the experiment, we will test the proposed learning scheme in the case study of face localization with the assumption that an image contains only one face region. Note that existing methods of face localization are generally based on learning face appearance [10, 12] that is not as flexible as the proposed scheme. Experiments are done on dataset YaleB [5]. YaleB contains 10 subjects, 9 poses, and 64 different illumination conditions in each pose. We manually extracted face region as the ground truth ROI.

Fig. 5 shows algorithmic ROIs associated with the optimal scales of images of three subjects, and Fig. 6 shows algorithmic ROIs associated with the optimal scales of images of different poses and illumination conditions. Observe that optimal scales under different illumination conditions are mostly consistent to each other, while optimal scales under different poses can differ significantly. This implies that appearance variation caused pose variations is an important factor in deciding optimal scales for ROI extraction. As a comparison, Fig.7 shows sampled results obtained by Viola-Jones face detection method [12]. We can observe that Viola Jones method fails when the face poses are relatively far away from the frontal pose.



$\sigma = 7$ $\sigma = 13$ $\sigma = 16$

Fig. 5. Algorithmic ROIs associated with optimal scales of images of different subjects.

With the structuring technique proposed in Section 2.2, we label all images with two classes, i.e., small-scale range and large-scale range. We use 2-fold validation to estimate the classification accuracy. The classification accuracy we obtained is 84%, which confirms the existence of substantial correlation between (Gaussian) scale spaces used in the context of localization and Gabor features used in the context of classification.

Furthermore, we test the running cost of ROI extraction with learning scale range, compared with the exhaustive search. Our test configuration consists of a computer of CPU Pentium 4 (3.40GHz) and Memory 4G with Matlab and C as the programming languages. Results are presented in Table 1, where the time unit is second.

sample step	with learning (sec)	without learning (sec)
1	31	42
2	21	34
3	9	18

Table 1. A comparison of running costs of ROI extraction with and without learning scale ranges.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a learning scheme for the classification of scale range, aiming to reduce the search space of

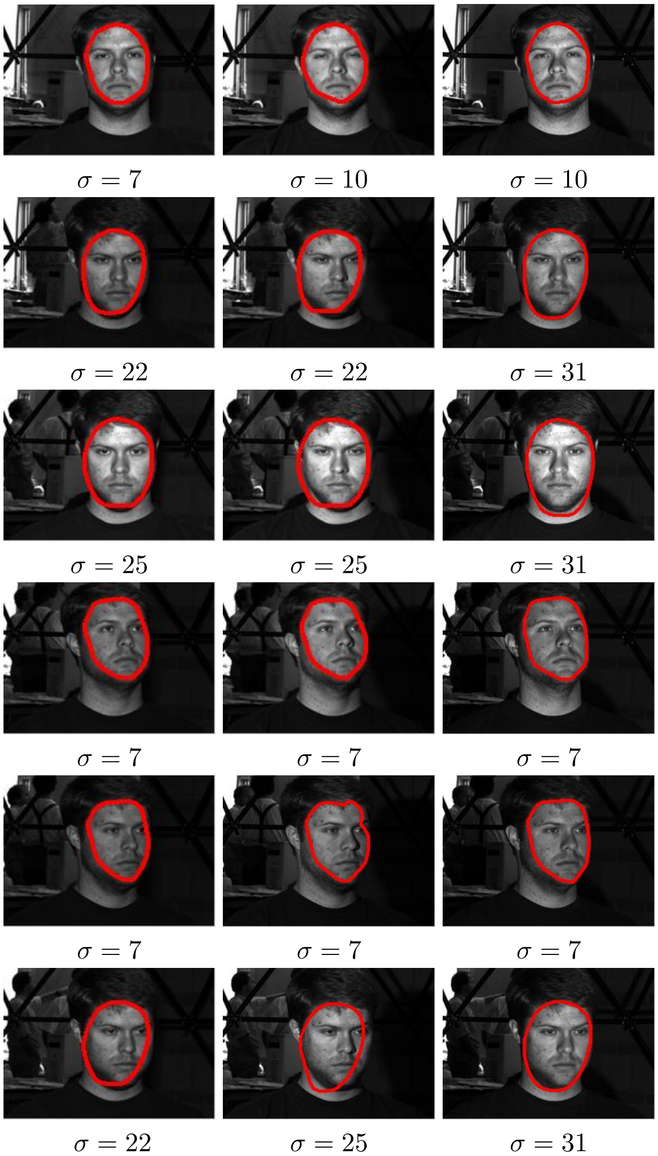


Fig. 6. Algorithmic ROIs associated with optimal scales of images of different poses and illuminations. A row represents a pose, and a column represents an illumination condition.

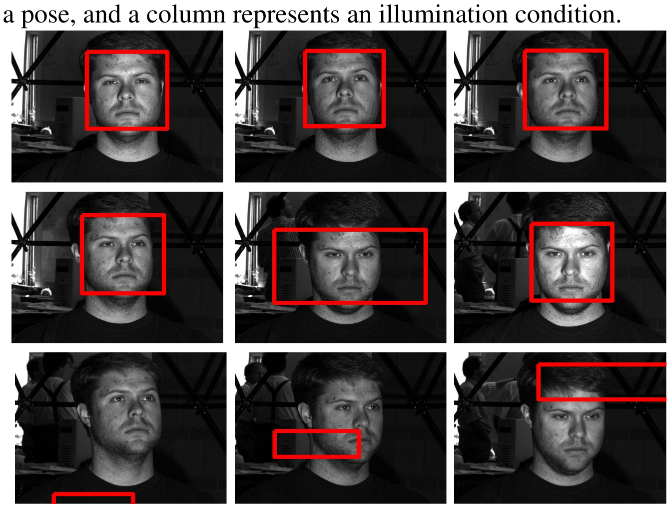


Fig. 7. Results obtained by Viola-Jones face detection method [12].

scales. Experimental results have shown the effectiveness of the proposed scheme. In the future, we plan to integrate the learning scheme of scales with other localizations methods, such as active contour, level sets, where weighting parameters are expected to be optimized too.

Acknowledgment: The work of Q. Li and Z. Bessinger was supported by National Science Foundation Grant IIS-1016668.

5. REFERENCES

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In *European Conference on Computer Vision (1)*, pages 45–58, 1996.

[2] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans on Acoustics, Speech, and Signal Processing*, 36(7):1169–1179, 1988.

[3] A. Dempster, N. Laird, and D. Rubin. Likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[4] C. Duanggate, B. Uyyanonvara, S. Makhanov, S. Barman, and T. Williamson. Object detection with feature stability over scale space. *J. Visual Communication and Image Representation*, 22(4):345–352, 2011.

[5] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[6] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[7] S. Nilufar, N. Ray, and H. Zhang. Object detection with DoG scale-space: A multiple kernel learning approach. *IEEE Trans Image Process, to appear*.

[8] B. Ommer and J. Malik. Multi-scale object detection by clustering lines. In *ICCV*, pages 484–491, 2009.

[9] S. Somayajula, A. Joshi, and R. Leahy. Mutual information based non-rigidmouse registration using a scale-space approach. In *IEEE International Symposium on Biomedical Imaging*, pages 1147–1150, 2008.

[10] K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.

[11] M. Toews, I. Wells, and M. William. A mutual-information scale-space for image feature detection and feature-based classification of volumetric brain images. In *IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), CVPR*.

[12] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[13] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition in still and video images: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.