Multi-scale Structural Saliency for Signature Detection

Guangyu Zhu[1], Yefeng Zheng[2], David Doermann[1], and Stefan Jaeger[1]

[1]Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742
[2]Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540

Abstract

Detecting and segmenting free-form objects from cluttered backgrounds is a challenging problem in computer vision. Signature detection in document images is one classic example and as of yet no reasonable solutions have been presented. In this paper, we propose a novel multi-scale approach to jointly detecting and segmenting signatures from documents with diverse layouts and complex backgrounds. Rather than focusing on local features that typically have large variations, our approach aims to capture the structural saliency of a signature by searching over multiple scales. This detection framework is general and computationally tractable. We present a saliency measure based on a signature production model that effectively quantifies the dynamic curvature of 2-D contour fragments. Our evaluation using large real world collections of handwritten and machine printed documents demonstrates the effectiveness of this joint detection and segmentation approach.

1. Introduction

Detecting free-form objects pose fundamental challenges in a number of aspects. First, detection needs to be robust in the presence of cluttered backgrounds. Second, non-rigid objects can have very large intra-class variations, making it almost impossible to model without overfitting the data. Third, the contours of many such complex objects are fragmented 2-D signals, so reliably recovering the ordering of points along contour fragments from off-line images is difficult in general. In addition, recognition and retrieval require well segmented objects from the detected regions, to minimize the effects of outliers during matching. Detecting signatures from documents is an example of one such difficult problem in which diverse layout structures, complex background, and noise make contour-based learning hard. Furthermore, the foreground content of documents generally includes a mixture of machine printed text, handwriting, diagrams, and other elements. Handwritten signature detection and segmentation is still an open research area and to our best knowledge, no comprehensive solutions have been presented in the literature.

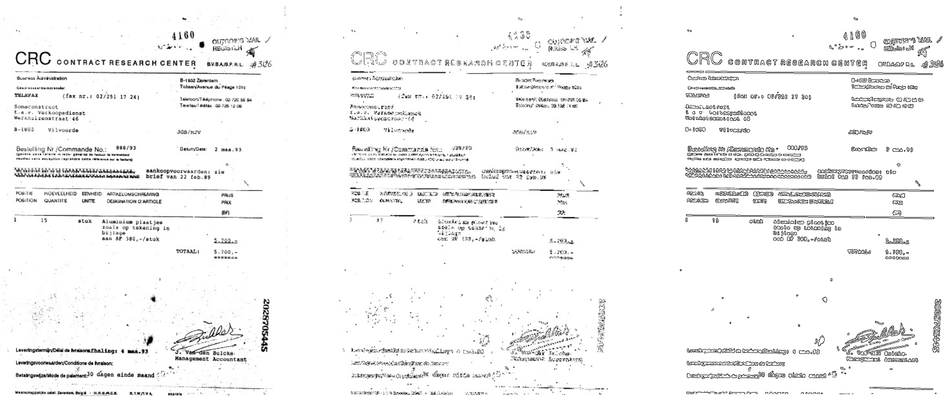
As signatures are a pervasive method of individual identification and document authentication, they provide an important form of indexing that enables exploration of large document repositories. Given a large collection of documents, searching for a specific signature is a highly effective way of retrieving documents authorized or authored by an individual. Such need arises frequently in the discovery phase of legal and intelligence investigations [11].

Prior research on off-line signatures has almost exclusively focused on signature verification and identification [18, 14, 10, 5] in the context of biometrics to perform authentication. For signature verification, the problem is to decide whether a sample signature is genuine or a forgery by comparing it with stored reference signatures. Signature identification is essentially a writer identification problem, whose objective is to find the author of a test signature given a database of signature exemplars from different signers. Most studies published to date assume that an almost perfect detection and segmentation is available [16].

Solving the problem of signature detection and segmentation is pivotal for signature-based document indexing and retrieval. Equally important is that a solution will also benefit off-line signature verification and identification in a range of domains. In addition, the ability to robustly detect signatures and extract them intact from volumes of documents is needed in many business and government applications.

In this paper, we propose a new multi-scale approach to detecting and extracting signatures from document images. Rather than viewing a signature as a collection of local features, we treat it as a global symbol that exhibits characteristic structural saliency. Computation in the proposed multi-scale framework for joint object detection and segmentation is carried out efficiently in a coarse-to-fine scheme on contour fragments. We employ a novel saliency measure based on a signature production model, which assumes two general degrees of freedom. The model enables us to capture the dynamic curvature in a signature without recovering its temporal information—a task shown to be difficult for unconstrained off-line handwriting due to structural changes [6]. The signature detection approach can also be applied to on-line handwritten notes collected on a PDA or Tablet PC, where the trajectories of the pen are readily available.

1-4244-1180-7/07/$25.00 ©2007 IEEE



(a) (b) (c)

Figure 1. The result of edge detection over different scales. (b) and (c) are edges computed from (a) at the scale of 2.0 and 4.0, respectively.

Our multi-scale detection framework is general and has the advantage that it does not embed explicit assumptions on local features of the object, such as the granulometric size distributions [18] and the sets of stroke-level features [8] for signatures. Therefore, it can be robust against many forms of variations that are challenges in shape-based object detection problems, and be generally applicable in spite of differences across languages. Computationally, the algorithm is tractable and highly parallelizable.

The structure of this paper is as follows: The next section reviews related work. In Section 3, we describe the multi-scale structural saliency detection approach. We introduce a saliency measure for capturing the dynamic curvature for signatures in Section 4. We discuss experimental results on real world English and Arabic document datasets in Section 5 and conclude in Section 6.

2. Related work

The *Saliency Network* by Sha'ashua and Ullman [20] considers boundary-based image segmentation and tries to jointly solve the two aspects of this problem iteratively, i.e. identifying salient structures and grouping contours. They defined a saliency function which monotonically increases with the length of the curve and monotonically decreases with the total squared curvature. To reduce the exponential search space, Sha'ashua and Ullman assumed the existence of a recurrent structure in the optimal solution, which they called *extensibility*, so that they can search the exponential space of possible curves in polynomial time by dynamic programming. However, greedily reducing the solution space by such a recurrent formulation involves hard decision at each step, and theoretically, a single mistake can result in the convergence to a wrong solution.

Alter and Basri [3] gave a comprehensive analysis of the saliency network and showed that its time complexity is $O(k^2N^2)$, where N is the total number of pixels and k is the number of neighboring elements considered in forming a locally connected network. They proved that the salient network has a few serious problems due to the extensibility assumption, and the convergence rates vary significantly depending on the object structure. These are difficult to overcome without fundamentally changing the computation.

There is a body of literature on contour grouping and contour-based learning in computer vision. Here we point out some of the work more related to ours, which includes Parent and Zucker's [15] work using relaxation methods, and Guy and Medioni's [9] work using voting patterns. Elder and Zucker [7] have developed a method for finding closed contours using chains of tangent vectors. Williams and Jacobs [23] and Williams and Thornber [24] discuss contour closure using stochastic completion fields. Shotton et al. [21] demonstrate a learning-based method for object detection using local contour-based features extracted from a single image scale. Such approach, however, is inherently restricted to rigid objects having 1-D contours, for which the ordering of points is explicit.

3. Multi-scale structural saliency

3.1. Theoretical framework

We consider identifying salient structure and grouping its structural components separately. There are clear motivations for decoupling these two tasks as opposed to solving them jointly. First, we have a much broader set of functions that can be used as measures of saliency. For object detection, choosing saliency measures that fit high-level knowledge of the object gives globally more meaningful results than jointly optimizing a fixed set of low-level vision constraints. Once the salient structures are identified, grouping becomes simpler with constraints like proximity and good continuation. Second, we can effectively formulate structural saliency across image scales, as apposed to single-scale approaches like the saliency network, for which a coarser representation will generally pose a harder problem. Multi-scale detection is important for non-rigid objects like signatures, whose contours can be severely broken due to poor ink condition and image degradations. Last, multi-scale saliency computation creates detection hypotheses at the natural scale where grouping among a set of connected components becomes structurally obvious. This provides a unified framework for object detection and segmentation, useful for both object recognition and retrieval.

From computational point of view, using connected components as the unit of computing saliency makes the computation tractable and highly parallelizable. Our serial implementation runs in $O(N)$, where N is the total number of edge points. This is significantly faster than the saliency network approach that has $O(k^2N^2)$ time complexity. We also incorporate global context to improve detection. The idea is to estimate the length and inter-line spacing of text lines, and use that information to locate the bottom region of the document, where signatures are more likely to appear. In our evaluation, we show results of signature detection run on whole documents, as well as by exploring global context.

3.2. Signature detection and segmentation

Unconstrained off-line signatures have large intra-class variations, and signatures from different individuals visually exhibit wider variability compared to other forms of handwriting because of their role in distinctively representing each person [8]. Another notable source of variation comes from intersession variability [18], the phenomenon that the signer cannot repeatedly write the same signature with exquisite precision over time. This factor greatly hinders the feasibility of posing signature detection and recognition as a straightforward pattern recognition problem. Large variations make classification approaches based on local features ineffective on real data. As shown in Fig. 2, local features, including size, aspect ratio, and spatial density, are not discriminative enough to separate signatures (red pluses) from non-signature objects (blue dots), on a groundtruthed collection of 1290 real-world documents.

In this section, we describe the structural saliency approach to signature detection that searches over range of scales $\mathcal{S} = \sigma_1, \sigma_2, \cdots, \sigma_n$. We select the initial scale σ_1 based on the resolution of the input image. We define the multi-scale structural saliency for a curve Γ as

$$\Phi(\Gamma) = \max_{\sigma_i \in \mathcal{S}} f(\Phi_{\sigma_i}(\Gamma_{\sigma_i}), \sigma_i), \qquad (1)$$

where $f\colon \mathbf{R}^2 \to \mathbf{R}$ is a function that normalizes the saliency over its scale, and Γ_{σ_i} is the obtained connected component corresponding to the curve at the scale σ_i. Using multiple scales for detection relaxes the requirement that the curve Γ be connected at a particular scale.

Detection at a particular scale σ_i proceeds in three steps. First, we convolve the image with a Gaussian kernel G_{σ_i}, re-sample it using the Lanczos filter [22] at the factor d_{σ_i},

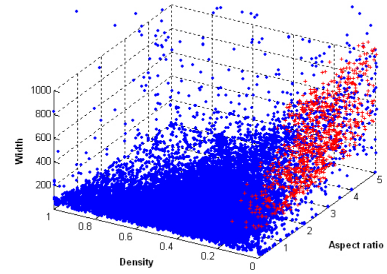


Figure 2. Local features, including size, aspect ratio, and spatial density, are not discriminative enough to separate non-signature objects (blue dots) from signatures (red pluses).

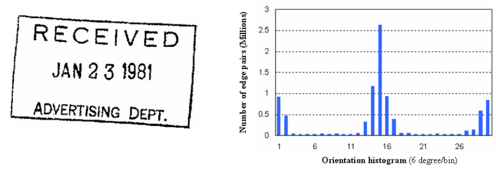


Figure 3. Example of histogram of orientation difference.

and compute its edges using the Canny edge detector [4]. This is effectively obtaining a coarse representation of the original image in which small gaps in the curve are bridged by smoothing followed by re-sampling (see Fig. 1).

Second, we form connected components on the edge image at scale σ_i, and compute the saliency of each component using the measure presented in Section 4, which characterizes the dynamic curvature in the curve. We define the saliency of a connected component $\Gamma^k_{\sigma_i}$ as the sum of saliency values computed from all pairs of edges on it. We also makes use of the distribution of pair-wise orientation difference, which is a global translation, rotation, and scale invariant shape descriptor. Using the histogram of pair-wise orientation difference, we can identify cases as shown in Fig. 3 to further reduce computation and false alarms.

Third, we identify the most salient curves and use a grouping strategy based on proximity and curvilinear constraints to obtain the rest of the signature parts within their neighborhood. Figs. 6 and 8 show examples of detected signatures from the Tobacco-800 and Maryland Arabic datasets, together with their saliency maps. The top three most salient components are shown in red, green and blue on the saliency maps.

Our joint detection and segmentation approach considers identifying the most cursive structure and grouping it with neighboring elements in two steps. By recognizing the most salient part of a signature, we effectively focus our attention on its neighborhood. Subsequently, a complete signature is segmented out from background by grouping salient neighboring structures.

Separating saliency detection from grouping significantly reduces the level of complexity. If we let the total number of edge points be N and the length of the largest connected component be L_c, the saliency computation is of order $O(NL_c)$. Since L_c is effectively bounded above by prior estimate of the signature dimensions and the range of searched scales n is limited, they can be considered as constants. The complexity in saliency computation is linear in N. Gaussian smoothing and forming connected component both require $O(N)$ time. The total complexity in the signature detection algorithm is therefore $O(N)$.

4. Measure of saliency for signatures

In this section, we consider the problem of recognizing the global saliency of a signature using dynamic curvature, without attempting to recover the temporal order among point sets on a 2-D curve. As shown in Fig. 4, among the infinite number of geometric curves that pass two given end points E_1 and E_2 on a signature, very few are realistic. This is because the wrist is highly constrained in the degrees of freedom when producing a signature. Furthermore, a signature segment rarely fits to a high-order polynomial, as shown by the dotted curve in Fig. 4.

We propose a signature production model that incorporates two general degrees of freedom in Cartesian coordinates. We assume that the pen moves in a cycloidal fashion with reference to a sequence of shifting virtual baselines. Local baseline changes as the wrist moves its position with respect to the document. Within a short curve segment, we assume that the baseline remains unchanged. In addition, the locus of the pen maintains a proportional distance from the local center point (*focus*) to the local baseline (*directrix*). This is equivalent to viewing handwriting approximated by a piece-wise concatenation of small elliptic segments. In a similar spirit, Saint-Marc et al. [19] have used quadratic B-splines to approximate complex-shaped contours. The model imposes an additional constraint that limits the group of second-order curves to smoother ellipses.

We model piece-wise segments of a signature by a family of second-order curves that satisfy constraints imposed by signature production. In addition, we incorporate the local gradient directions at the two end points, which can be viewed as soft constraints on the segment of the curve imposed by the global structure of the signature instance. In the Cartesian coordinate system, the family of a quadratic equation in two variables x and y is always a conic section. Fig. 5 shows how the orientations of the gradients at the two edge points greatly limit the inference on local curve segment to a family of conics, under the second-order signature production model.

We can formalize this intuition geometrically. For a pair of edge points E_1 at (x_1, y_1) and E_2 at (x_2, y_2), we obtain estimates of their local gradients $N_1(p_1, q_1)$ and $N_2(p_2, q_2)$ during edge detection. For definiteness, we suppose both E_1 and E_2 point into the angular section between the tangent lines containing the other point, as shown in Fig. 5.

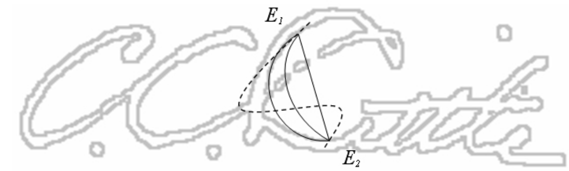


Figure 4. Among the large number of geometric curves can pass the two end points E_1 and E_2 on a signature, few are realistic.

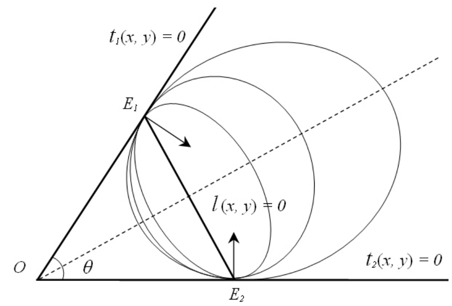


Figure 5. Conic sections inferred by a pair of edge points.

$$p_1(x_2 - x_1) + q_1(y_2 - y_1) > 0 \quad \text{and}$$
$$p_2(x_1 - x_2) + q_2(y_1 - y_2) > 0 \tag{2}$$

The two tangent lines at E_1 and E_2 are normal to their local gradients and are given by

$$t_1(x, y) \equiv p_1(x - x_1) + q_1(y - y_1) = 0 \tag{3}$$

and

$$t_2(x, y) \equiv p_2(x - x_2) + q_2(y - y_2) = 0. \tag{4}$$

The straight line $l(x, y)$ that passes through E_1 and E_2 can be written as

$$l(x, y) \equiv \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0. \tag{5}$$

Note that $t_1(x, y)$, $t_2(x, y)$ and $l(x, y)$ are all first-order linear functions in x and y. This family of second-order curves that are bounded within the angular section between $t_1(x, y)$ and $t_2(x, y)$ can be expressed in the canonical form in Equation (6), where parameters a, b, c, f, g, h are first-order linear functions in λ and the parameter set (x_1, y_1), (x_2, y_2), (p_1, q_1), (p_2, q_2). Interested readers can refer to Maxwell's classic geometry text [13] for details.

$$\begin{aligned} C(x, y) &\equiv l^2(x, y) - \lambda t_1(x, y) t_2(x, y) = 0 \\ &= ax^2 + 2hxy + by^2 + 2gx + 2fy + c = c \end{aligned} \tag{6}$$

Given a parameter set (x_1, y_1), (x_2, y_2), (p_1, q_1), (p_2, q_2), which is equivalent to fixing the set of three straight lines $t_1(x, y)$, $t_2(x, y)$ and $l(x, y)$, it is interesting to see how

Figure 6. Examples of detected signatures from the Tobacco-800 dataset, together with their saliency maps. The top three most salient parts are shown in red, green, and blue, respectively.

monotonic increase in λ for $\lambda \in [0, +\infty)$ affects the curvature within the bounded segment of the quadratic curve defined by (6). When $\lambda = 0$, Equation (6) degenerates into the straight line $l(x, y)$ and the total squared curvature of the bounded segment is zero. When λ monotonically increases from 0 within certain range ($0 < \lambda < \lambda_0$), the curve segment bounded by E_1 and E_2 exhibits more and more curvature. This is because the second-order curves given by (6) for $\lambda \in [0, \lambda_0)$ are ellipses with monotonically increasing eccentricities. As $\lambda \rightarrow \lambda_0$, the center of the ellipse recedes to infinity, so that the ellipse tends to a parabola at $\lambda = \lambda_0$. When $\lambda \geq \lambda_0$, the conic of (6) becomes a hyperbola. Eventually as $\lambda \rightarrow +\infty$, the hyperbola degenerates into the two intersected straight lines $t_1(x, y)$ and $t_2(x, y)$. We can show that λ_0 is given as

$$\lambda_0 = \frac{4[p_1(x_2 - x_1) + q_1(y_2 - y_1)]}{(p_1 q_2 - p_2 q_1)} \times \frac{[p_2(x_1 - x_2) + q_2(y_1 - y_2)]}{(p_1 q_2 - p_2 q_1)}. \tag{7}$$

The value of λ_0 provides a second-order estimate of the dynamic curvature given the parameter set (x_1, y_1), (x_2, y_2), $(p_1,\ q_1)$, $(p_2,\ q_2)$ that fits the signature production model. We use λ_0 as the saliency value $\Lambda_{\sigma_i}(E_i, E_j)$ for a pair of points at scale σ_i. The cursive saliency of a connected component $\Gamma^k_{\sigma_i}$ at scale σ_i is the sum of saliency defined by Equation (7) computed from all the pairs of points on it.

$$\Phi_{\sigma_i}(\Gamma^k_{\sigma_i}) = \sum_{E_i, E_j \in \Gamma^k_{\sigma_i}} \Lambda_{\sigma_i}(E_i, E_j). \tag{8}$$

It is straightforward to show that the right hand side of (7) is $4a^2$, where $a = |OE_1| = |OE_2|$. This result allows us to normalize saliency over scale, whereas the relationship between the scale and the saliency measures reported in [24] are largely unclear. Obviously, the proposed measure of saliency is translation and rotation invariant as it only uses local gradient directions.

The analysis so far considers the continuous case. To account for the discretization effect on the image grid, we impose two conditions. First, the absolute values of the two functions on the left hand side of Equation (2) must be strictly large than ϵ. Second, the denominator term in (7) must be strictly large than ϵ. In our implementation, we use $\epsilon = 0.1$. For robustness, we weight the saliency contribution by the gradient magnitude of the weaker edge.

5. Experiments

5.1. Datasets

To evaluate the structural saliency approach for signature detection on multiple languages, we used two large collections of real world documents—*Tobacco-800* dataset and the *University of Maryland Arabic* dataset. Tobacco-800 is a public subset of the IIT CDIP Test Collection [1, 11], based on 42 million pages of documents (in 7 million multi-page TIFF images) obtained from UCSF [2] and released by tobacco companies under the Master Settlement Agreement. Tobacco-800 is a realistic dataset for document analysis and retrieval as these documents were collected and scanned using a wide variety of equipment over time. In addition, a significant percentage of Tobacco-800 are consecutively numbered multi-page business documents, making it a valuable testbed for various content-based document retrieval approaches. The Maryland Arabic dataset consists mainly of Arabic handwritten business documents. Using public datasets gives more realistic evaluation in contrast to

Table 1. Summary of the English and Arabic evaluation datasets.

	Tobacco-800	*Maryland Arabic*
Document Types	Printed/handwritten	Mostly handwritten
Total Pages	1290	169
Resolution (in DPI)	150–300	200
Labeled Signatures	900	149

common published evaluations using self collected datasets that captures much less variations. Typical dimensions of documents range from 1200×1600 to 2500×3200 pixels in Tobacco 800 and 1700×1200 to 2000×2600 pixels in Maryland Arabic.

The groundtruth of signatures were manually labeled in rectangular boxes using our developed Java editor [17]. Whenever possible, we also label the identity of the signer by reconciling the document context. This enables quantitative evaluation on signature retrieval, where the identities of the signers are required. Since the number of signatures vary significantly across documents, we assume no prior knowledge on the distribution of signatures per document. In our evaluation, we use all the documents in the Tobacco-800 and Maryland Arabic datasets.

5.2. Evaluation

We focused on two aspects in our evaluation. First, we use the detection probability P_D and false-alarm probability P_F as metrics. P_D and P_F represent the two degree of freedom in a binary hypothesis test and they do not involve a prior probabilities of the hypothesis. To factor in the "*quality*" of the detection, we consider a signature *correctly detected* and *complete* if the detected region overlaps with more than 75% of the labeled signature region. We declare a *false alarm* if the detected region does not overlap with more than 25% of any labeled signature region.

Fig. 7 shows the ROC curves on the Tobacco-800 and Maryland Arabic datasets. Fisher classifier using size, aspect ratio, and spatial density features serve as a baseline for comparison, with all other procedures remaining the same in the comparison experiment. We use two scale levels in multi-scale detection experiments. Parameters involved in obtaining the ROC curves, including the cutoff threshold in saliency and estimated signature dimensions, are tuned on 10 documents. We use the following approach to compute each operating point on an ROC curve. After we compute the saliency of each signature candidate, we store it with the internal zone representation of the candidate. We apply a reasonably low global decision threshold for detection and sort the ranked list of detected candidates from the entire test set by their saliencies. To plot a new point on the ROC curve, we move down the ranked list by one and look at the portion of the ranked list from its top to the current position, which is equivalent to gradually lowering the global

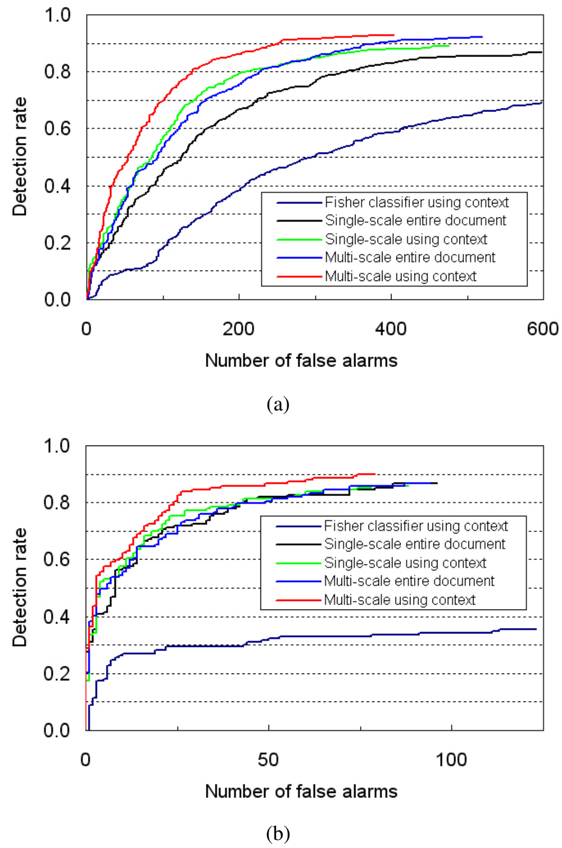


Figure 7. ROC curves for (a) Tobacco-800 dataset and (b) Maryland Arabic dataset.

decision threshold. The entire sets of ROC curves computed by this scheme as shown in Fig. 7 are highly densely packed and include every operating point.

Multi-scale saliency approach gives best overall detection performance on both English and Arabic datasets. Using document context, our multi-scale signature detector achieves 92.8% and 86.6% detection rates for the Tobacco-800 and Maryland Arabic datasets, at 0.3 false-positives per image (FPPI). Encouragingly, the advantage of multi-scale approach becomes more obvious on a more diverse dataset, like Tobacco-800. Exploring global context is more effective on machine printed documents as geometric relationships among text lines are more uniform.

Second, we test how discriminative is our proposed saliency measure in capturing the global cursive pattern embedded in signatures. The handwritten Maryland Arabic dataset serves better for this purpose, because variations among local features including size, is not discriminative, as evident from the poor performance of Fisher classifier.

Figs. 6 and 8 show samples of detected signatures from Tobacco-800 and Maryland Arabic datasets, together with their saliency maps. We show the top three most salient parts in red, green, and blue, respectively. In our experiment, a cursive structure is normally more than an order

Figure 8. Examples of detected signatures from the Maryland Arabic dataset, together with their saliency maps. The top three most salient parts are shown in red, green, and blue, respectively.

of magnitude more salient than printed text of the same dimensions. However, we did find a few instances of printed text among false alarms that obtain saliencies comparable to signatures because of their highly cursive fonts, as shown in Fig. 9(a). A limitation of our proposed method is that the detected and segmented signature may contain a few touching printed characters when signatures overlap with very strong background. Nevertheless, the quality of segmented output by structural saliency is considerably better.

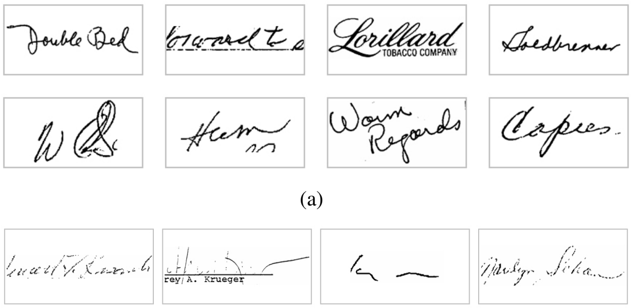
For better interpretation of the overall detection performance, we summarize key evaluation statistics. On Tobacco-800, 848 signatures out of the 900 labeled signatures are correctly detected, by the multi-scale saliency approach using document context in Fig. 7(a). Among correctly detected signatures, 83.3% are complete. Their mean percentage area overlap with the groundtruth is 86.8% with a standard deviation of 11.5%. As shown in Figs. 6 and 8, the quality of detected signatures is comparable to manually cropped versions. This demonstrates that using connected components give extracted signatures of impressive quality, and it does not necessarily limit the detection probability when used in a multi-scale approach. In fact, these figures are close to the machine printed text word segmentation performance level from leading commercial OCR product on Tobacco-800 documents. The results on the Maryland Arabic dataset are also very encouraging as the collection consists mainly of unconstrained handwriting in complex layouts and backgrounds.

5.3. Discussion

On the saliency maps, an edge detector generates two parallel contour segments from a stroke since both of them are local maxima in gradient magnitude. A ridge detector can reduce the level of saliency computation and give more compact segmentation output, since it generates only one contour response for each stroke. However, a ridge detector [12] performs much worse in signature detection in our experiments. This is because the Canny edge detector provides good localization that guarantees accurate estimation of local gradient directions, even under significant amount of degradation. Another factor in favor of using edges is that saliency computation takes only a small portion of the total computation, as compared to obtaining the coarse-scale representations of the image and running edge detection.

Some examples of false positives from the Tobacco-800 set are shown in Fig. 9(a), which include cases of handwriting. The classification between signature and handwriting is sometimes not well posed by considering only shape. Highly cursive handwritten words may not have any obvious visual differences from signatures, as illustrated by handwriting shown in Fig. 9(a). Using document context could not effectively resolve such intricacies because they are primarily handwritten annotations that are semantically associated with the printed content in the document. To some extent, semantics should be exploited to solve the ambiguity in this case.

Fig. 9(b) shows examples of false negatives in detection. These missed signatures are so severely broken, that a step edge operator like Canny could not form reasonable contour fragments, even at a coarse scale. As shown on most signatures, however, using multiple scales for detection partially overcome the limitations of connected-components-based approach by relaxing the requirement that the contour fragment be well connected at a particular scale. This improvement is more clearly observed on the Tobacco-800 dataset, which contains more highly degraded images at low resolution.



(a)

(b)

Figure 9. Examples of (a) false alarms and (b) missed signatures from the Tobacco-800 dataset.

6. Conclusion

In this paper, we propose a novel signature detection approach based on the view that object detection can be a process that aims to capture the characteristic global structural saliency of the object over multiple scales. This is different from the common object detection framework that focuses on sets of local properties of the object. The results on signature detection on multi-language datasets show that our approach is very effective on real document collections that have large variations. One advantage using multi-scale saliency approach for joint detection and segmentation is that it provides a general framework for which detection and segmentation degrades gracefully as the problem becomes more challenging. In addition, detected and segmented output looks both structurally and perceptually meaningful.

Acknowledgements

This work was supported by the US Department of Defense under contract MDA-9040-2C-0406 and in part by a subcontract from an ARDA Challenge Grant on Complex Document Image Processing 4-31809.

References

[1] The IIT complex document image processing (CDIP) test collection, http://ir.iit.edu/projects/CDIP.html, 2006.

[2] The legacy tobacco document library (LTDL) at UCSF, http://legacy.library.ucsf.edu/, 2006.

[3] T. Alter and R. Basri. Extracting salient curves from images: an analysis of the saliency network. *Int. J. Computer Vision*, 27(1):51–69, 1998.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 8(6):679–697, 1986.

[5] S. Chen and S. Srihari. Use of exterior contours and word shape in off-line signature verification. In *Proc. Int. Conf. Document Analysis and Recog.*, pages 1280–1284, 2005.

[6] D. Doermann and A. Rosenfeld. Recovery of temporal information from static images of handwriting. *Int. J. Computer Vision*, 15(1-2):143–164, 1995.

[7] J. Elder and S. Zucker. The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Res.*, 33(7):981–991, 1993.

[8] K. Guo. *Forgery detection by local correspondence*. PhD thesis, University of Maryland, College Park, 2000.

[9] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *Int. J. Computer Vision*, 20(1-2):113–133, 1996.

[10] M. Kalera, S. Srihari, and A. Xu. Off-line signature verification and identification using distance statistics. *Int. J. Pattern Recog. and Artificial Intell.*, 18(7):1339–1360, 2004.

[11] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J. Heard. Building a test collection for complex document information processing. *SIGIR*, pages 665–666, 2006.

[12] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Int. J. Computer Vision*, 30(2):77–116, 1996.

[13] E. Maxwell. *The Methods of Plane Projective Geometry Based on the Use of General Homogeneous Coordinates*. Cambridge Univ. Press, 1946.

[14] M. Munich and P. Perona. Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification. In *Proc. Int. Conf. Computer Vision*, pages 108–115, 1999.

[15] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(8):823–839, 1989.

[16] R. Plamondon and S. Srihari. On-line and off-line handwriting recognition: a comprehensive survey. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(1):63–84, 2000.

[17] M. Roth, S. Jaeger, and D. Doermann. GEDI: Ground truth editor and document interface. *Summit on Arabic and Chinese Handwriting Recognition*, 2006.

[18] R. Sabourin, G. Genest, and F. Preteux. Off-line signature verification by local granulometric size distributions. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(9):976–988, 1997.

[19] P. Saint-Marc, H. Rom, and G. Medioni. B-spline contour representation and symmetry detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 15(11):1191–1197, 1993.

[20] A. Sha'ashua and S. Ullman. Structural saliency: the detection of globally salient structures using a locally connected network. In *Proc. Int. Conf. Computer Vision*, pages 321–327, 1988.

[21] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. In *Proc. Int. Conf. Computer Vision*, pages 503–510, 2005.

[22] K. Turkowski. Filters for common resampling tasks. *Graphics Gems I*, pages 147–165, 1990.

[23] L. Williams and D. Jacobs. Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Computation*, 9:849–870, 1997.

[24] L. Williams and K. Thornber. A comparison of measures for detecting natural shapes in cluttered backgrounds. *Int. J. Computer Vision*, 33(2-3):81–96, 2000.