

Распознавание символов на искаженных изображениях с использованием морфологической и фазовой фильтраций

Х. Диаз-Эскобар*, В.И. Кобер**, В.Н. Карнаухов**

* Департамент Компьютерных наук, Центр Научных исследований и высшего образования, Энсенада, 22860

** Институт проблем передачи информации, Российская академия наук, Москва, 127051, Россия

Поступила в редколлегию 22.5.2015

Аннотация—В настоящее время активно разрабатываются новые алгоритмы для оптического распознавания символов для мобильных устройств. Большинство этих алгоритмов работают хорошо с документами, оцифрованными в идеальных условиях. При использовании мобильных устройств изображения документов подвержены различным искажениям, таким как неравномерное и плохое освещение, геометрические искажения, плохое разрешение, сенсорные шумы. В данной работе мы предлагаем надежный метод распознавания контекстуально-несвязанных символов на искаженных изображениях документов с использованием банка адаптивных морфологических и фазовых фильтров. С помощью компьютерного моделирования проведено сравнение результатов работы традиционных и предлагаемого подходов для обнаружения и классификации символов латинского алфавита на искаженных изображениях.

КЛЮЧЕВЫЕ СЛОВА: распознавание символов, фазовая фильтрация, морфологическая корреляция.

ВВЕДЕНИЕ

В последние годы популярность мобильных устройств (гаджетов) стремительно растет. Бурный рост количества мобильных устройств вызвал большой рост количества различных приложений для них, в том числе тех, которые используют оптическое распознавание символов. В настоящее время многие алгоритмы распознавания, предлагаемые и используемые в таких популярных коммерческих системах, как ABBYY (<http://www.abbyy.com>), Tesseract [1], OCRopus [2], и др., обычно работают с “идеальными” и специально отформатированными данными, обеспечивающими высокую производительность. Изображения, получаемые в реальных мобильных устройствах, обычно регистрируются в реальных, а не в идеальных условиях съемки. По этой причине качество изображений часто оказывается низким из-за наличия геометрических искажений, неоднородности освещения, низкого разрешения, шума регистратора и т.д. Эти эффекты существенно влияют на производительность систем оптического распознавания символов [3]–[5].

Одной из основных проблем в системах распознавания является неоднородность освещения. Большинство систем распознавания символов используют бинарные изображения, теряя важную информацию в процессе бинаризации. Кроме того, корректная сегментация символов становится практически невозможной. Системы, основанные на выделении признаков, такие как метод К-средних, метод главных компонент, метод опорных векторов, нейронные сети, метод глубокого обучения и т.д., часто используются для распознавания символов. Тем не менее, такие системы зачастую упускают из анализа некоторую важную информацию при расчете признаков символов и, следовательно, дают ошибки классификации. Кроме того, нейронные

сети и метод глубокого обучения требуют сложной многоуровневой структуры и нуждаются в длительном обучении с использованием наборов образцов [6]–[14].

В этой работе мы используем банк адаптивных нелинейных композитных фильтров и фазовых фильтров для обнаружения и классификации символов. Предлагаемый подход основан на пороговой декомпозиции, инвариантной к освещению, морфологической корреляции и модификации фазового фильтра. Адаптивные композитные фильтры синтезируются, используя информацию из набора обучающих изображений. Заданный уровень дискриминации для этих фильтров обычно достигается на стадии итерационного обучения. В статье также представлены результаты компьютерного моделирования, полученные с помощью предлагаемой системы.

Статья организована следующим образом: в разделе 1 описана предлагаемая система оптического распознавания символов, в разделе 2 представлены полученные результаты компьютерного моделирования и раздел 3 суммирует наши выводы.

1. ПРЕДЛАГАЕМАЯ СИСТЕМА ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ

Предлагаемая система работает в два этапа: предобработка и распознавание.

Предобработка

Вначале мы используем нерезкое маскирование для повышения локального контраста между передним планом и фоном. Затем, для подавления смеси аддитивного и импульсного шума во входной сцене используется ранговая фильтрация [15]–[19]. Наконец, для коррекции освещенности входная сцена локально нормализуется в пределах скользящего окна W малого размера, тогда сигнал в окне можно считать равномерно освещенным. Обозначим исходную сцену – $S(\vec{x})$, эталонный объект – $T(\vec{x})$, тогда откорректированная сцена $\hat{S}(\vec{x})$ вычисляется как

$$\hat{S}(\vec{x}) = a(\vec{x})S(\vec{x}) + b(\vec{x}), \quad (1)$$

где $\vec{x} = (x_1, x_2)$ – координаты во входной сцене, $a(\vec{x})$ и $b(\vec{x})$ – локальные коэффициенты нормализации, учитывающие локальную освещенность и смещение сигнала в точке с координатами $\vec{x}$. Коэффициенты, оптимальные относительно средней квадратичной ошибки, задаются следующими выражениями [20]:

$$a(\vec{x}) = \frac{\sum_{\vec{n} \in W} T(\vec{n}) \cdot \hat{S}(\vec{n} + \vec{x}) - |W| \cdot \bar{T} \cdot \bar{S}(\vec{x})}{\sum_{\vec{n} \in W} [\hat{S}(\vec{n} + \vec{x})]^2 - |W| \cdot [\bar{S}(\vec{x})]^2}, \quad (2)$$

$$b(\vec{x}) = \bar{T} - a(\vec{x})\bar{S}(\vec{x}), \quad (3)$$

где $\bar{T}$ и $\bar{S}(\vec{x})$ – средние значения сигнала объекта и сигнала сцены в локальном окне W в точке с координатами $\vec{x}$, соответственно.

Распознавание

Второй этап обработки – распознавание. Цель этого этапа состоит в том, чтобы распознать символы и вычислить координаты каждого символа на изображении документа. Чтобы добиться этого, мы предлагаем использовать нелинейный фильтр синтетических дискриминантных функций (NSDF) и морфологическую корреляцию.

В соответствии с концепцией пороговой декомпозиции [21] полутоновое изображение $S(\vec{x})$, имеющее Q уровней квантования, можно представить в виде суммы бинарных срезов $\{S^q(\vec{x}), q =$

$1, \dots, Q-1$ следующим образом:

$$S(\vec{x}) = \sum_{q=1}^{Q-1} S^q(\vec{x}). \quad (4)$$

Пусть $\{T_i(\vec{x}), i = 1, \dots, N\}$ – N полутоновых обучающих изображений, принадлежащих к истинному классу объектов распознавания. NSDF фильтр может быть записан следующим образом [20],[22]:

$$H_{NSDF}(\vec{x}) = \sum_{q=1}^{Q-1} \bigcap_{i=1}^N T_i^q(\vec{x}), \quad (5)$$

где $\bigcap_{i=1}^N$ – пересечение между N бинарными плоскостями, а $\{T_i^q(\vec{x}), q = 1, \dots, Q-1; i = 1, \dots, N\}$ – бинарные изображения, полученные с помощью пороговой декомпозиции из соответствующих обучающих изображений $\{T_i(\vec{x}), i = 1, \dots, N\}$.

С другой стороны, известно, что при определенных условиях минимизация средней абсолютной ошибки $E(\vec{x})$ между входным изображением $S(\vec{x})$ и сдвинутой версией объекта $T(\vec{x})$ в точке с координатами $\vec{x}$:

$$E(\vec{x}) = \sum_{\vec{n} \in W} |S(\vec{x} + \vec{n}) - T(\vec{x})| = \sum_{\vec{n} \in W} S(\vec{x} + \vec{n}) + \sum_{\vec{n} \in W} T(\vec{x}) - 2 \sum_{\vec{n} \in W} \min(S(\vec{x} + \vec{n}), T(\vec{n})), \quad (6)$$

эквивалентна максимизации последнего члена уравнения (6) – нелинейной корреляции между ними [23] и, в соответствии с пороговой декомпозицией, нелинейная корреляция может быть вычислена как

$$C(\vec{x}) = \sum_{\vec{n} \in W} \sum_{q=1}^{Q-1} [S^q(\vec{x} + \vec{n}) \cap T^q(\vec{n})], \quad (7)$$

где $\{S^q(\vec{x}), q = 1, \dots, Q-1\}$ и $\{T^q(\vec{x}), q = 1, \dots, Q-1\}$ являются бинарными изображениями, полученными из полутоновых изображений входной сцены и объекта, соответственно, и $\cap$ – пересечение между двумя бинарными плоскостями.

Наконец, выполняется нелинейная корреляция и нормализуется коэффициентом $\frac{u}{t}$. Здесь u – предварительно определенное значение выхода коррелятора для каждого обучающего изображения и $t = \sum_{\vec{n} \in W} H_{NSDF}(\vec{n})$, где $H_{NSDF}(\vec{n})$ – синтезированный фильтр синтетических дискриминантных функций из N полутоновых обучающих изображений объекта. Окрестность W берется как пересечение между областями определения всех обучающих изображений [20], [22].

Существует компромисс между количеством обучающих изображений, используемых для синтеза фильтра, и качеством работы фильтра. Для решения этой задачи, предлагается использовать многоуровневый банк фильтров. На каждом уровне символы делятся на k классов, и для каждого класса синтезируется нелинейный дискриминантный фильтр. Фильтры синтезируются таким образом, чтобы в местах локализации символов на выходе системы получались резкие корреляционные пики, при этом в фоновой области выходные значения системы должны быть близкими к нулю. Для синтеза фильтров использовался адаптивный итерационный алгоритм [24]. Этот алгоритм позволяет подавить ложные корреляционные пики для данного класса фонов.

Дискриминационная способность (DC) используется для оценки способности фильтра к распознаванию заданного объекта среди других объектов. DC формально определяется следующим образом [25]

$$DC = 1 - \frac{|C_{\max}^b|^2}{|C_{\max}^t|^2}, \quad (8)$$

где $C_{\max}^b$ и $C_{\max}^t$ – максимальные значения в корреляционной плоскости по области фона и цели, соответственно.

Для повышения эффективности распознавания будем использовать также дуальный объект [26]. Пусть $\{T_i(\vec{x}), i = 1, \dots, N\}$ – набор N полутоновых обучающих изображений, используемых для синтеза фильтра $H_{NSDF}(\vec{n})$. Вычислим дуальный фильтр $H_{NSDF_{dual}}(\vec{n})$, используя пороговую декомпозицию,

$$H_{NSDF_{dual}}(\vec{n}) = \sum_{q=1}^{Q-1} \bigcap_{i=1}^N T_{dual_i}^q(\vec{n}) \quad \text{при} \quad T_{dual_i}^q(\vec{n}) = 1 - T_i^q(\vec{n}), \quad \vec{n} \in W, \quad q = 1, \dots, Q-1. \quad (9)$$

Зададим дуальный объект следующим образом:

$$H_{2L_{dual}}(\vec{n}) = \begin{cases} 255, & \vec{n} \in RS \\ 0, & \text{в противном случае,} \end{cases}$$

где RS – область определения H_{NSDF} . Далее, вычислим нелинейные корреляции, используя уравнение (7) для $H_{2L}(\vec{n})$ и $H_{2L_{dual}}(\vec{n})$. Известно, что при некоторых предположениях [26], максимизация разницы между двумя морфологическими корреляциями эквивалентна минимизации $E(\vec{x})$ в уравнении (6).

Так как некоторые символы алфавита очень похожи, то эффективность их распознавания с помощью нелинейных дискриминантных фильтров невысока. В этих случаях предлагается использовать другой подход, основанный на маскировании нулем спектральных компонент фазовых фильтров [27]. Предположим, что входная сцена является вещественной и содержит объект распознавания $t(\vec{x})$ и ложный объект $b(\vec{x})$. Уравнение (8) может быть записано в следующем виде:

$$DC = 1 - \frac{\left| 2 \int_{-\infty}^{+\infty} \int_0^{+\infty} |B(\vec{u})| \cos[\Delta\varphi(\vec{u})] d\vec{u} \right|^2}{\left| 2 \int_{-\infty}^{+\infty} \int_0^{+\infty} |T(\vec{u})| d\vec{u} \right|^2}, \quad (10)$$

где $B(\vec{u})$ и $T(\vec{u})$ – преобразования Фурье $b(\vec{x})$ и $t(\vec{x})$, соответственно. Подставляя предварительно заданное значение DC, например, $(1 - \varepsilon^2)$ в уравнение (10) и, используя формулу $(|a|^2 - |b|^2) = (a - b)(a + b)$, мы приходим к следующему результату:

$$\left(\int_{-\infty}^{+\infty} \int_0^{+\infty} |B(\vec{u})| \cos[\Delta\varphi(\vec{u})] - \varepsilon |T(\vec{u})| d\vec{u} \right) \times \left(\int_{-\infty}^{+\infty} \int_0^{+\infty} |B(\vec{u})| \cos[\Delta\varphi(\vec{u})] + \varepsilon |T(\vec{u})| d\vec{u} \right) = 0. \quad (11)$$

Теперь, задача состоит в достижении равенства нулю произведения двух выражений. Алгоритм состоит из двух шагов: на первом шаге отсчеты первого сомножителя уравнения (11) упорядочиваются в вариационный ряд; затем, выбирается минимальное количество отсчетов в вариационном ряду, которые будут блокированы нулями так, чтобы удовлетворить условию

равенства в уравнении (11). Если, например, результат вычислений выражения в первых скобках больше нуля, то обнуляем элементы вариационного ряда, начиная с наибольшего значения. На втором шаге, описанная процедура, повторяется для элементов второго сомножителя уравнения (11).

Наконец, для того чтобы учесть геометрические искажения объектов распознавания, синтезируется банк нелинейных дискриминационных фильтров для каждого класса объектов, используя множество известных представлений символов. Обнаружение может быть выполнено путем вычисления корреляции входной сцены и каждого фильтра из банка фильтров. Далее, вычисляется дискриминационная способность DC в каждой корреляционной плоскости, и плоскость с наивысшим значением DC выбирается, как выход всей системы. Координаты корреляционного пика в выбранной плоскости используются в качестве оценки положения объекта на сцене [28].

2. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Тестовые изображения содержали символы, искаженные с помощью таких преобразований как поворот в диапазоне $[-20, 20]$ градусов с шагом 1 градус, растяжение с коэффициентом 0.8, горизонтальный сдвиг в диапазоне $[-0.6, 0.6]$ с шагом 0.05 и изменение масштаба в диапазоне $[0.8, 1.3]$ с шагом 0.1. Средний размер символов – 31×31 пикселей. Обучающие изображения объектов, использованные для синтеза каждого фильтра в банке фильтров, были повергнуты повороту в диапазоне $[-15, 15]$ градусов с шагом 3 градуса, растянуты с коэффициентом 0.8 (вдоль осей x и y), сдвинуты в направлении оси x в диапазоне $[-0.6, 0.6]$ с шагом 0.05 и подвергались масштабированию в диапазоне $[0.8, 1.1]$ с шагом 0.1.

Были проведены эксперименты с различными параметрами шума и освещенности при одновременном изменении положения символов на входной сцене. Тестовые сцены были одновременно искажены и аддитивным шумом со среднеквадратичным отклонением σ , равным 5, 10, 15, и импульсным шумом с вероятностью $p = 5\%$. Неравномерное освещение моделировалось с помощью Ламбертовой модели [29]

$$d(\vec{x}) = \cos \left\{ \frac{\pi}{2} - \arctan \left[\frac{\rho}{\cos(\Theta)} [(\rho \tan(\Theta) \cos(\varphi) - x_1)^2 + (\rho \tan(\Theta) \cos(\varphi) - x_2)^2]^{-\frac{1}{2}} \right] \right\}, \quad (12)$$

где параметры имели следующие значения: $\Theta = 65$, $\varphi = 60$, а $\rho = 70$, $\rho = 50$, $\rho = 40$ и $\rho = 30$.

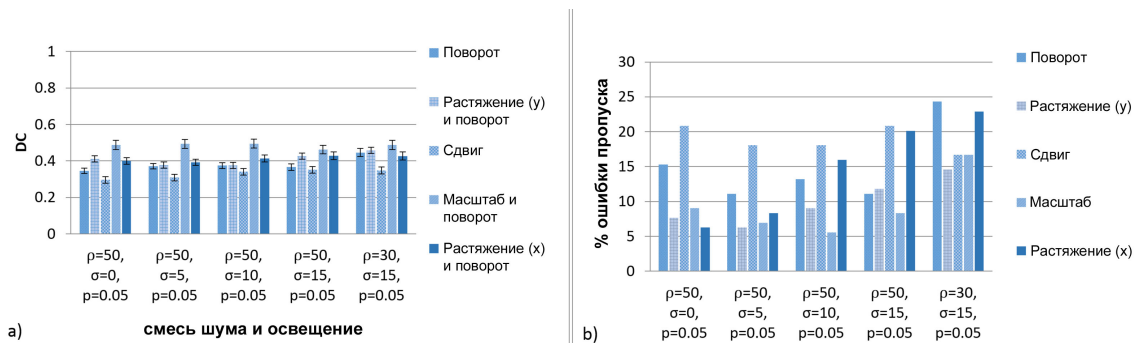


Рис. 1. Распознавания двух схожих искаженных объектов: а) DC с достоверностью 95%, б) ошибки пропуска.

Минимальное значение DC равно 0.36. На Рис. 1(а) показаны результаты распознавания двух схожих объектов с помощью предложенной фазовой фильтрации: $Class1 = \{p\}$, $Class2 = \{b\}$. Процент ложных обнаружений – не более 5%, ошибки пропуска показаны на Рис.1(б).

Максимальное значение DC равно 0.73. На Рис. 2(а) показана эффективность распознавания с помощью фильтра H_{NSDF} , использованного для распознавания и классификации следующих классов символов:

$$Class1 = \{b, p, d, g, q, o, e, c, s, a, m, h, n, r, u\}, \quad Class2 = \{r, t, f, z, k, x, y, w, v, l, j, i\},$$

ошибки пропуска показаны на Рис. 2(б). На Рис. 3(а) представлена сравнительная характеристика работы по распознаванию предлагаемой системы и коммерческой системы ABBYY на реальном изображении, представленном на Рис. 3(б).

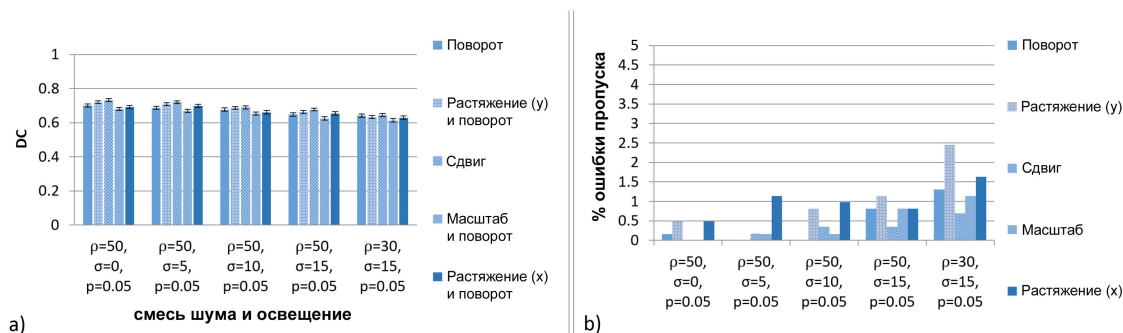


Рис. 2. Распознавания двух классов символов: а) DC с достоверностью 95%, б) ошибки пропуска.

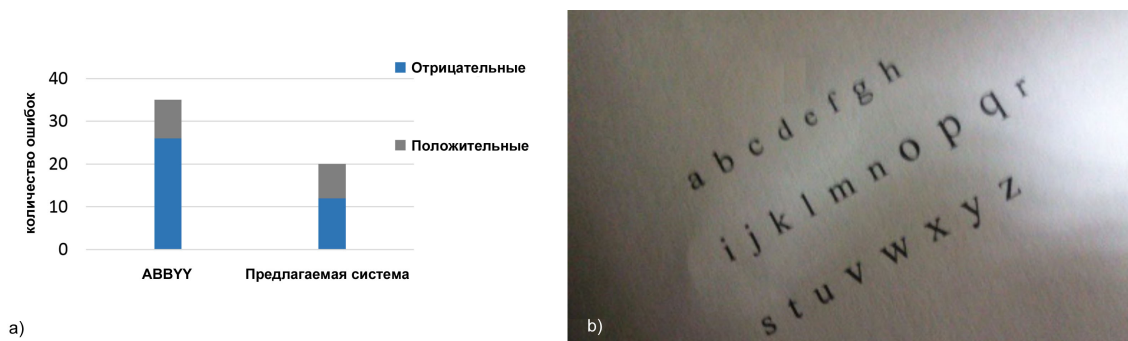


Рис. 3. Сравнительная характеристика работы по распознаванию предлагаемой системы и коммерческой системы ABBYY – (а) на реальном изображении – (б).

ЗАКЛЮЧЕНИЕ

В данной статье предложена система распознавания символов латинского алфавита с использованием банка адаптивных нелинейных дискриминационных фильтров на искаженных изображениях документов. Предлагаемые методы дают хорошие результаты с точки зрения распознавания и скорости обработки при наличии шума и неравномерного освещения на сценах с геометрическими искажениями. Как и ожидалось, коммерческая система ABBYY показывает высокую эффективность распознавания в тех случаях, когда изображение имеет хорошее качество, однако при низкой или неравномерной освещенности сцены ее эффективность резко уменьшается. В дальнейшей работе предлагается использовать адаптивные методы сегментации и контекстуальное распознавание полных слов. Для увеличения скорости обработки пред-

полагается реализовать предложенную систему с использованием графических процессоров и технологий параллельного программирования.

СПИСОК ЛИТЕРАТУРЫ

1. Breuel T.M. Exact solutions to total variation regularization problem. *In IS&T/SPIE Elec. Imag., Int. Soc. Opt. and Phot.*, 68150-68150, 2008.
2. Patel C., Patel A. and Patel D. Optical character recognition by open source OCR tool tesseract: A case study. *IJCA*, 2012, vol. 55, No. 10, pp. 50-56.
3. Holley R. How good can it get? Analyzing and improving OCR accuracy in large scale historic newspaper digitization programs. *D-Lib Magazine*, 2009, vol. 15, pp. 3-4.
4. Ozarslan S., and Eren P. E. Text recognition and correction for automated data collection by mobile devices. *In IS&T/SPIE Elec. Imag., Int. Soc. Opt. and Phot.*, 2014, pp. 902706-902706.
5. Esser D., Muthmann K. and Schuster D. Information extraction efficiency of business documents captured with smartphones and tablets. *In Proc. of ACM symp. on Doc. Eng.*, 2013, pp. 111-114.
6. Lund W. B., Kennard D. J. and Ringger E. K. Combining multiple thresholding binarization values to improve OCR output. *In IS&T/SPIE Elec. Imag., Int. Soc. for Opt. and Phot.*, 2013, pp. 6580-86580 .
7. Gatos B., Pratikakis I. and Perantonis S. J. Adaptive degraded document image binarization. *Patt. Rec.*, 2006, vol. 39, No. 3, pp. 317-327.
8. Ulges A., Lampert C. H. and Breuel T. M. Document image dewarping using robust estimation of curled text lines. *In DAR, Proc. 8th Int. Conf. on IEEE*, 2005, pp. 1001-1005.
9. Zhang H., Zhao K., Song Y. Z. and Guo J. Text extraction from natural scene image: A survey. *Neuro-computing*, 2013, vol. 122, pp. 310-323.
10. Ma D. and Agam G. A super resolution framework for low resolution document image OCR. *In IS&T/SPIE Elec. Imag., Int. Soc. for Opt. and Phot.*, 2006, pp. 86580-86580.
11. Matei O., Pop P. C. and Velea H. Optical character recognition in real environments using neural networks and k-nearest neighbor. *Applied intelligence, Springer*, 2013, vol. 39, No. 4, pp. 739-748.
12. Salehpour M. and Behrad A. Cluster based weighted SVM for the recognition of Farsi handwritten digits. *In NNA, Elec. Eng., 10th Symp., IEEE*, 2010, pp. 219-239.
13. Kir B., Oz C. and Gulbag A. The Application of optical character recognition for mobile device via artificial neural networks with negative correlation learning algorithm. *In ICECCO, Int. Conf. on IEEE*, 2013, pp. 220-223.
14. LeCun Y., Bottou L., Bengio Y. and Haffner P. Gradient-based learning applied to document recognition. *In Proc. of the IEEE*, 1998, vol. 86, No. 11, pp. 2278-2324.
15. Kober V., Mozerov M., Alvarez-Borrego J. and Ovseyevich I.A. Adaptive rank-order correlations. *Pattern Recognition and Image Analysis*, 2009, vol. 14, No. 1, pp. 33-39.
16. Kober V. Robust and Efficient Algorithm of Image Enhancement. *IEEE Trans. on Cons. Elec*, 2009, vol. 52, No. 2, pp. 655-659 .
17. Kober V., Mozerov M., Alvarez-Borrego J. Nonlinear filters with spatially connected neighborhoods. *Opt. Eng.*, 2001, vol. 40, No. 6, pp. 971-983 .
18. Kober V., Mozerov M., Alvarez-Borrego J. and Ovseyevich I.A. Rank image processing using spatially adaptive neighborhoods. *Pattern Recognition and Image Analysis*, 2001, vol. 11, No. 3, pp. 542-552 .
19. Kober V., Mozerov M., Alvarez-Borrego J. and Ovseyevich I.A.. Unsharp masking using rank-order filters with spatially adaptive neighborhoods. *Pattern Recognition and Image Analysis*, 2002, vol. 12, No. 1, pp. 46-56.

20. Martinez-Diaz S., and Kober V. Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition. *Opt. Eng.*, 2008, vol. 47, No. 6, pp. 067201–067201 .
21. Fitch J., Coyle E.J. and Gallagher Jr. N.C. Median filtering by threshold decomposition. *ASSP, Int. Conf. on. IEEE Trans.*, 1984, vol. 32, No. 6, pp. 1183–1188.
22. Kober V., Diaz-Martinez S., Karnaukhov V and Ovseyevich I.A. Distortion invariant pattern recognition with local correlations. *Pattern Recognition and Image Analysis*, 2011, vol. 21, No. 2, pp. 188–191 .
23. Maragos P. Morphological correlation and mean absolute error criteria. *ASSP, Int. Conf. on. IEEE*, 1568–15721.
24. Gonzalez-Fraga J. A., Kober V. and Alvarez-Borrego J. Adaptive synthetic discriminant function filters for pattern recognition. *Opt. Eng.*, 2006, vol. 45, No. 5, pp. 057005–057005.
25. Yaroslavsky L.P. The theory of optimal methods for localization of objects in pictures. *In Proc. In Opt., Elsevier*, 1993, vol. 32, No. 1, pp. 145–201.
26. Maragos P. Optimal morphological approaches to image matching and object detection. *In Proc. Int. Conf. on CV.*, 1998, pp. 695–699 .
27. Aguilar-Gonzalez P.M., Kober V. and Diaz-Ramirez V. H.. Adaptive composite filters for pattern recognition in nonoverlapping scenes using noisy training images. *Pattern Recognition Letters*, 2014, vol. 41, No. 1, pp. 83–92 .
28. Kober V. I. and Ovseyevich I.A. Phase-Only filter with improved filter efficiency and correlation discrimination. *Pattern Recognition and Image Analysis*, 2000, vol. 10, No. 4, pp. 514–519 .
29. Diaz-Ramirez, V. H., Picos, K. and Kober, V. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*, 2014, vol. 32, No. 1, pp. 32–43.

Character recognition in degraded document images using morphological and phase-only filtering

Diaz-Escobar J., Kober V., Karnaukhov V.

Nowadays applications of optical character recognition (OCR) for mobile devices have been actively developing. Most of these algorithms work well with clean and well-formed data, but images of mobile devices usually are captured in unfavorable conditions. Therefore, the quality of document images often is poor owing to various distortions such as nonhomogeneous illumination, perspective distortions, low resolution, sensor noise, etc. In this work we propose a method for reliable recognition of characters in degraded document images using a bank of adaptive morphological and phase-only filters. With the help of computer simulation the performance of the proposed method is presented and compared with that of common approaches for detection and classification of characters of the Latin alphabet in distorted document images.

KEYWORDS: variational functional, total variation, exact solution, regularization, signal restoration.