

# Kombinasi MSER Dan SURF Dalam Mendeteksi Teks Pada Gambar Natural

Kgs Achmad Siddik<sup>1</sup>, Yohannes<sup>2</sup>

STMIK GI MDP

Jl. Rajawali No 14 Palembang, Indonesia

{siddiqachmad, yohannesmasterous}@mdp.ac.id

Saparudin

Fakultas Ilmu Komputer Universitas Sriwijaya

Palembang, Indonesia

saparudin1204@yahoo.com

**Abstrak**— Deteksi teks pada sebuah citra merupakan hal yang harus dilakukan sebelum melakukan pengenalan karakter. Proses deteksi teks berperan penting dalam perolehan informasi pada sebuah citra. Pada penelitian ini dilakukan deteksi teks secara otomatis menggunakan Maximally Stable Extremal Regions (MSER) dan Speeded-up Robust Features (SURF). Kombinasi MSER dan SURF digunakan untuk menemukan wilayah kandidat teks pada sebuah citra. Kemudian edge detection juga dilakukan untuk melakukan validasi terhadap wilayah teks yang telah ditemukan. Berikutnya, metode Stroke Width Transform (SWT) digunakan untuk membedakan wilayah citra teks dan non-teks. Dataset yang digunakan adalah ICDAR 2003. Hasil pengujian menunjukkan bahwa hasil kombinasi MSER dan SURF dapat mendeteksi teks pada natural image secara lebih baik dibandingkan dengan MSER saja.

**Kata Kunci** — *Deteksi teks, Maximally Stable Extremal Region, Speeded-up Robust Features, Stroke Width Transform*

## I. PENDAHULUAN

Deteksi teks pada gambar natural merupakan tantangan tersendiri karena citra dari gambar memiliki *background* yang kompleks. Gambar natural juga mempunyai beragam objek dalam sebuah gambar pemandangan sehingga wilayah teks yang akan dideteksi pada gambar tersebut menjadi tantangan seperti ukuran, jenis, dan bentuk teks. Maximally Stable *External Region* (MSER) adalah metode yang telah banyak digunakan untuk mendeteksi area teks pada gambar natural secara *Unsupervised* seperti yang dilakukan penelitian sebelumnya [1][2][3][4]. Selain menggunakan MSER deteksi teks juga dapat dilakukan dengan *Histogram Of Oriented Gradients* (HOG) [5] telah menghasilkan akurasi yang baik akan tetapi deteksi teks tidak dilakukan secara langsung atau *unsupervised*. Dalam upaya meningkatkan akurasi MSER juga telah dikombinasi dengan beberapa metode salah satunya dengan menggunakan *higher-order correlation clustering*

(HOCC) agar dapat meningkatkan akurasi dari teks yang didapatkan pada region MSER[6].

Metode Speed Up Robust Feature (SURF) sendiri telah banyak digunakan untuk mencocokkan kedua gambar secara otomatis [7][8] dan telah menghasilkan akurasi pencocokan gambar dengan sangat baik, bahkan metode MSER dan SURF telah digunakan untuk mencocokkan kedua gambar secara otomatis[9] dan telah menghasilkan akurasi yang sangat baik dari penelitian sebelumnya. Oleh karena itu metode yang diajukan pada penelitian ini adalah mengkombinasi Maximally Stable External Region (MSER) dengan Speed Up Robust Feature (SURF) yang tujuan dari mengkombinasi kedua metode ini untuk meningkatkan akurasi dari penelitian yang telah dilakukan sebelumnya. Pada bagian II akan dijelaskan studi literature dari penelitian sebelumnya dalam mendeteksi teks pada gambar natural, pada bagian III dijelaskan bagaimana alur dari metode yang diajukan pada penelitian ini dan pada bagian. Dataset yang digunakan pada penelitian ini menggunakan dataset ICDAR 2013 yang dimana data ini telah digunakan juga oleh penelitian sebelumnya.

## II. STUDI LITERATUR

### 2.18. Deteksi Region: MSER

Fitur pada MSER secara umum tergantung dengan batas-batas wilayah dari citra yang dideteksi. Struktur algoritma MSER dan algoritma DAS pada dasarnya identik, tetapi struktur output dari kedua algoritma tersebut sangat berbeda. Dalam deteksi algoritma MSER satu set *threshold segment* pada gambar biner MSER dapat ditentukan dengan membandingkan sesuai bagian dari gambar tersebut diproses. MSER dibagi menjadi dua kategori berdasarkan kecerahan pada gambar natural dan dilambangkan dengan MSERs+ dan MSERs-.

Dalam menentukan hasil dari daerah *External* pencacahan dapat dilakukan dengan cara berikut:

*Prosiding*  
**ANNUAL RESEARCH SEMINAR 2016**

6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

1. *Pixel* diurutkan menurut intensitas.
2. Setelah pengurutan piksel ditempatkan dalam gambar dan daftar komponen yang terhubung dan daerah MSER dipertahankan dengan melakukan union antar daerah.
3. Selama proses pencacahan daerah masing-masing komponen yang terhubung sebagai fungsi intensitas disimpan.
4. Diantara banyak region yang dihasilkan, yang secara maksimal stabil adalah nilai dari region yang sesuai dengan *thresholds*.
5. Daerah MSER adalah bagian dari gambar dimana *binarization* stabil selama rentang dari nilai region masih didalam *thresholds*

#### 2.19. SURF

Secara umum algoritma *Speeded Up Robust Feature* (SURF) memiliki empat langkah utama: *Hessian Matrix* berbasis point deteksi yang masih pada ruang skala, *Interest Point Localization*, *orientation segment* dan Deskriptor berdasarkan jumlah respon dari Haar wavelet.

- (1) *Hessian Matrix* berbasis point deteksi dalam skala:

Matiks Hessian digunakan sebagai deskriminan untuk mencari nilai maksimal. Untuk setiap point  $X=(x,y)$  didalam gambar  $I$ , Hessian matriks  $H(X, \sigma)$  didalam  $X$  pada skala  $\sigma$  didefinisikan seperti pada rumus dibawa ini.

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix}$$

dimana  $L_{xx}(X, \sigma)$  adalah konvolusi dari turunan kedua Gaussian dari citra  $I$  pada titik  $X$  dan sama untuk  $L_{xy}(X, \sigma)$  dan  $L_{yy}(X, \sigma)$ . Kemudian aproksimasi dari hasil konvolusi dinyatakan sebagai  $D_{xx}$ ,  $D_{yy}$ , dan  $D_{xy}$  sehingga determinan dari matriks Hessian adalah  $\det(H_{approx}) = D_{xx} D_{yy} - (D_{xy})^2$ .

Kedua, proses lokalisasi interest point dimulai dengan melakukan penetapan terhadap nilai ambang untuk mendeteksi titik ekstrim dari matriks Hessian. Kemudian sebuah non maksimum dalam ukuran  $3 \times 3 \times 3$  neighborhood diterapkan untuk menghasilkan titik ekstrim. Hanya titik dengan nilai yang lebih besar dari 26 neighborhood lainnya yang dipilih sebagai feature point.

Ketiga, penetapan orientasi dimulai dari menghitung Haar wavelet dalam arah  $x$  dan  $y$ . Hasilnya dibobotkan dengan Gaussian yang berpusat

pada titik tersebut. Kemudian orientasi yang dominan ditentukan dengan menghitung jumlah semua hasil dalam orientasi ukuran masing-masing neighborhood. Keempat, penentuan deskriptor berdasarkan jumlah dari Haar wavelet yang kemudian dilakukan esktraksi dari descriptor tersebut dengan membangun wilayah persegi yang berpusat pada interest point dan orientasi di wilayah terseleksi sebelumnya.

#### 2.20. Deteksi Tepi

Deteksi tepi digunakan untuk mendeteksi tepi objek dari sebuah citra. Deteksi tepi sering digunakan dalam deteksi atau pengenalan sebuah objek. Deteksi tepi juga digunakan pada tahap preprocessing. Dalam hal ini deteksi tepi digunakan untuk mendeteksi tepi dari sebuah teks. Untuk mendeteksi tepi secara sempurna dari sebuah teks, metode penekanan non maxima digunakan sebagai hasil dalam satu pixel tepian yang luas sebagai output. Selanjutnya, MSER dikombinasikan dengan deteksi tepi untuk meningkatkan atau lebih tepatnya dikatakan hasil tepi yang ditingkatkan. Tepi yang diperoleh digunakan sebagai batas untuk daerah MSER dan informasi diluar batas tersebut akan dihapus. Pada kombinasi ini, arah gradien dari setiap pixel tepi disesuaikan sehingga dapat memberikan peningkatan representasi teks dalam kondisi blur.

#### 2.21. Stroke Width Transform (SWT)

[10]SWT diterapkan dalam rangka untuk menemukan kandidat karakter dalam teks. Blok yang terdeteksi yang mengandung calon karakter dikelompokkan secara bersama-sama dan blok yang tidak terdeteksi akan dihapus. Metode ini memungkinkan dapat diterapkan untuk beberapa bahasa dan jenis font.

SWT adalah sebuah operator citra lokal yang menghitung lebar stroke yang paling mungkin mengandung pixel. Stroke didefinisikan sebagai bagian kontinu dari citra yang membentuk sebuah band dengan lebar yang hampir konstan. Lebar yang hampir konstan merupakan salah satu fitur yang membedakan teks dari unsur-unsur lain pada citra, yang dapat dimanfaatkan untuk memulihkan daerah yang mungkin berisi teks. SWT melakukan integrasi informasi secara bottom-up, di mana ia menggabungkan piksel lebar stroke yang sama ke dalam komponen yang lebih besar. Output dari SWT adalah citra dengan ukuran yang sama dengan ukuran citra input di mana setiap elemen mengandung lebar stroke terkait dengan pixel.

2.22. Penelitian Terkait

[9] mengusulkan kombinasi Maximally Stable Extremal Regions (MSER) dan Speeded Up Robust Features (SURF) untuk image matching. Fitur SURF ke dikelompokkan ke dalam wilayah yang telah diekstrak oleh MSER dengan mengukur kesamaan fitur yang dimiliki. Kombinasi fitur yang didapatkan mampu meningkatkan image matching secara signifikan.

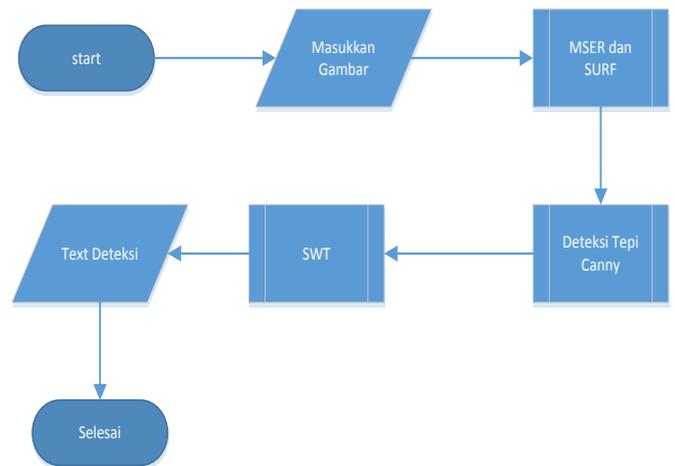
2.23. Kebaruan Dan Kontribusi

MSER telah digunakan untuk membangun wilayah kandidat teks pada sebuah citra di beberapa penelitian. Di sisi lain, kombinasi MSER dan SURF telah digunakan untuk image matching dengan hasil yang lebih baik dibandingkan dengan SURF.

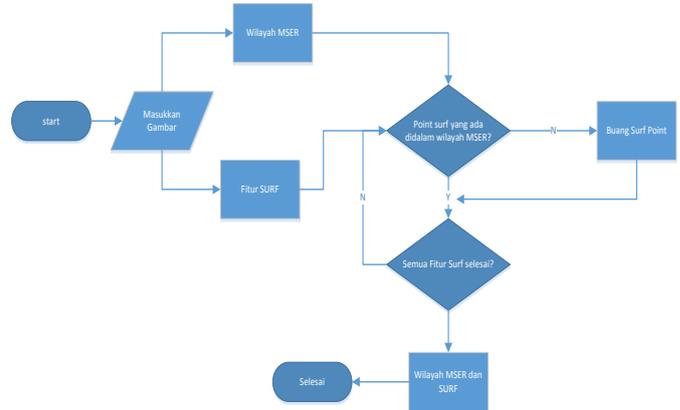
Kombinasi MSER dan SURF tersebut digunakan untuk membangun wilayah yang mengandung kandidat karakter pada sebuah citra (natural image). Kemudian dilakukan edge detection untuk melakukan validasi terhadap wilayah teks yang telah ditemukan dan selanjutnya digunakan SWT sebagai penentu dalam membedakan wilayah citra yang mengandung teks dan non-teks

III. METODE YANG DIUSULKAN

Penelitian ini mengusulkan deteksi teks secara otomatis menggunakan MSER dan SURF tahapan yang dilakukan untuk deteksi teks pada gambar natural dapat dilakukan dimulai dari input citra lalu dilakukan deteksi fitur MSER dan SURF. Kemudian dilakukan deteksi tepi dengan menggunakan canny agar dapat mendeteksi tepi dari sebuah teks, selanjutnya hasil wilayah yang dibangun oleh MSER dan SURF dikombinasi dengan deteksi tepi untuk meningkatkan hasil tepi kandidat teks. Tepi yang diperoleh menjadi batas untuk daerah MSER dan SURF. Stroke Width Transform (SWT) diterapkan untuk menentukan kandidat karakter dalam teks, blok yang terdeteksi dikelompokkan secara bersama-sama dan membentuk bounding box sebagai hasil akhir. Alur dari metode yang diusulkan dapat dilihat pada gambar berikut.



Gambar 1. Alur metode yang diusulkan



Gambar 2. Kombinasi MSER dan SURF

Gambar 2 merupakan alur dari kombinasi algoritma MSER dan SURF yang dimana citra awal diekstrak fitur dengan menggunakan MSER untuk membangun wilayah yang berpotensi mengandung teks. Citra awal juga di ekstrak fitur dengan menggunakan SURF. Wilayah yang telah dibangun oleh MSER akan dicocokkan terhadap fitur SURF kemudian dicari kesamaan lokasi point dari wilayah SURF dengan MSER. Wilayah kombinasi yang dibangun berdasarkan lokasi kesamaan tersebut membentuk wilayah baru sebagai wilayah kandidat yang mengandung teks.

*Prosiding*  
**ANNUAL RESEARCH SEMINAR 2016**  
*6 Desember 2016, Vol 2 No. 1*

---

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

---

REFERENSI

- [1] K. Venkateswarlu and S. M. Velaga, "Text Detection On Scene Images Using MSER," *Int. J. Res. Comput. Commun. Technol.*, vol. 4, no. 7, pp. 452–456, 2015.
- [2] A. Tabassum, "Text Detection using MSER and Stroke Width Transform," 2015.
- [3] R. Turki, M. Ben Ralima, and A. M. Alimi, "Scene Text Detection Images With Pyramid Image and MSER Enhanced," pp. 301–306.
- [4] M. Sungl, B. Jun, H. Ch, and D. Kiml, "Scene Text Detection with Robust Character Candidate Extraction Method," pp. 426–430, 2015.
- [5] A. Pise, "Text Detection and Recognition in Natural Scene Images," pp. 1068–1072, 2014.
- [6] L. Kang, Y. Li, and D. Doermann, "Orientation Robust Text Line Detection in Natural Images," pp. 4034–4041, 2014.
- [7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [8] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF : Speeded Up Robust Features."
- [9] L. Tao, X. Jing, S. Sun, H. Huang, N. Chen, and Y. Lu, "Combining SURF with MSER for Image Matching," pp. 286–290, 2013.
- [10] B. Epshtein, "Detecting Text in Natural Scenes with Stroke Width Transform," pp. 2963–2970, 2010.