

Detektor MSER dan OCR pada Pembacaan Rambu Petunjuk Jalan

Rahman Arifuddin^{1*}, Subairi², Elta Sonalitha³, Basitha Febrinda Hidayatullail⁴, Resi Dwi Jayanti Kartika Sari⁵,
Muhammad Darul Fatoni⁶

^{1,2,3,4,5,6}Teknik Elektro, Fakultas Teknik Universitas Merdeka Malang

Email: ¹rahman.arifuddin@unmer.ac.id, ²subairi@unmer.ac.id, ³elta.sonalitha@unmer.ac.id, ⁴basitha@unmer.ac.id,
⁵resi.sari@unmer.ac.id, ⁶darulfatoni5@gmail.com

*Penulis Korespondensi

Abstract — Text recognition on road guide signs remains a significant challenge in Digital Image Processing, particularly under varying illumination conditions, image distortions, and differences in font size. This paper explores the use of Maximally Stable Extremal Regions (MSER), a feature detection method, for text segmentation in road sign images, followed by the application of Optical Character Recognition (OCR) to identify the detected characters. MSER is employed to identify and extract stable regions within an image, enabling effective separation of text from background or non-text elements. It has proven efficient in addressing visual challenges such as shadows, reflections, and low image quality. Once text regions are successfully segmented, OCR is utilized to convert the extracted image-based text into machine-readable digital text, allowing the system to recognize various font types and text orientations more accurately. Experimental results show that the MSER method achieves a precision of 85.7% and a recall of 85.7%, resulting in an F1-score of 85.7%. Meanwhile, the OCR system demonstrates an average text recognition accuracy of 91.6%. Detection errors primarily occur under low-light conditions and skewed image capture angles, which affect the stability of regions during MSER extraction and subsequently impact OCR performance. This approach offers an effective solution for text segmentation and recognition in dynamic environments, such as road signs influenced by perspective variations and poor lighting conditions. Furthermore, it contributes to the development of automated road sign recognition systems that can be integrated into autonomous vehicles and image-processing-based applications.

Intisari— Pengenalan teks pada rambu petunjuk jalan merupakan tantangan utama dalam pengolahan citra digital, terutama ketika menghadapi kondisi pencahayaan yang bervariasi, distorsi gambar, dan perbedaan ukuran font pada rambu. Artikel ini membahas penggunaan *Maximally Stable Extremal Regions* (MSER), sebuah metode deteksi fitur, untuk segmentasi teks pada gambar rambu jalan, yang kemudian dilanjutkan dengan penerapan *Optical Character Recognition* (OCR) untuk mengenali karakter-karakter yang terdeteksi. MSER berfungsi untuk mengidentifikasi dan mengekstrak wilayah-wilayah stabil dalam gambar sehingga membantu memisahkan teks dari latar belakang atau elemen non-teks, serta terbukti efisien dalam mengatasi tantangan visual akibat bayangan, pantulan cahaya, dan kualitas gambar yang rendah. Setelah teks berhasil disegmentasi, OCR digunakan untuk mengonversi teks dalam gambar menjadi teks digital yang dapat diproses oleh komputer, sehingga memungkinkan sistem mengenali berbagai jenis *font* dan orientasi teks dengan lebih akurat. Berdasarkan hasil pengujian, metode MSER mampu mendeteksi area teks dengan nilai *precision* sebesar 85,7% dan *recall* sebesar 85,7%, sehingga menghasilkan *F1-score* sebesar 85,7%, sementara sistem OCR menunjukkan rata-rata akurasi pembacaan teks sebesar 91,6%. Kesalahan deteksi terutama terjadi pada kondisi pencahayaan rendah serta sudut pengambilan citra yang miring, yang mempengaruhi kestabilan region pada proses ekstraksi MSER dan berdampak pada hasil

pembacaan OCR. Pendekatan ini menawarkan solusi efektif untuk segmentasi dan pengenalan teks dalam kondisi lingkungan yang dinamis, seperti pada rambu lalu lintas yang sering terpengaruh sudut pandang atau pencahayaan buruk, serta berkontribusi pada pengembangan sistem pengenalan rambu otomatis yang dapat diintegrasikan dengan kendaraan otonom dan aplikasi berbasis pengolahan citra.

Kata Kunci— MSER, OCR, Segmentasi Teks, Rambu Petunjuk Jalan, Pengolahan Citra, Kendaraan Otonom.

I. PENDAHULUAN

Pengenalan teks pada gambar merupakan salah satu tantangan utama dalam bidang pengolahan citra digital, khususnya ketika teks tersebut terdapat pada objek yang memiliki latar belakang kompleks atau terdistorsi, seperti pada rambu petunjuk jalan. Rambu lalu lintas memainkan peran yang sangat penting dalam sistem transportasi, memberikan informasi vital bagi pengemudi dan pejalan kaki. Oleh karena itu, pengolahan citra yang efektif untuk mendeteksi dan membaca teks pada rambu jalan sangat diperlukan, terutama untuk aplikasi dalam kendaraan otonom, sistem navigasi otomatis, atau untuk meningkatkan sistem keselamatan jalan.

Dalam konteks transportasi modern, rambu lalu lintas berfungsi sebagai media komunikasi visual antara sistem transportasi dan pengguna jalan. Rambu tidak hanya memberikan peringatan atau larangan, tetapi juga menyampaikan informasi arah, lokasi, jarak, dan panduan yang esensial bagi pengemudi. Pada sistem kendaraan otonom atau semi-otonom, kemampuan untuk secara otomatis mengenali dan menafsirkan teks pada rambu jalan merupakan salah satu kemampuan kunci yang menentukan keselamatan dan efisiensi navigasi. Karena itu, bidang pengolahan citra yang berfokus pada deteksi dan pengenalan teks dari citra rambu jalan telah menjadi topik riset yang menarik dan relevan secara global.

Dalam konteks Indonesia, penerapan sistem pengenalan teks pada rambu jalan masih memiliki peluang riset yang sangat luas. Salah satu tantangan unik di Indonesia adalah variasi desain dan kondisi rambu yang sangat beragam. Ada rambu dengan latar biru, hijau, kuning, bahkan putih, dengan teks dalam berbagai bahasa atau singkatan. Kondisi lingkungan tropis juga menimbulkan masalah tambahan seperti refleksi cahaya kuat, kelembapan tinggi, dan permukaan rambu yang sering berdebu. Oleh karena itu, model deteksi dan OCR yang dikembangkan di negara lain belum tentu langsung dapat diterapkan secara optimal di sini tanpa penyesuaian dataset dan parameter

Namun, mengenali teks dalam gambar tidaklah semudah tampak di permukaan. Tantangan utama muncul dari kondisi

lingkungan yang tidak terkendali: pencahayaan yang berubah-ubah, sudut pandang kamera yang tidak sejajar, bayangan dari objek sekitar, hingga distorsi akibat gerakan kendaraan. Semua faktor ini dapat menyebabkan teks pada rambu menjadi buram, terdistorsi, atau tidak kontras terhadap latar belakangnya. Kondisi semacam ini membuat proses deteksi teks konvensional sering gagal mengidentifikasi karakter secara akurat. Oleh karena itu, dibutuhkan metode segmentasi yang robust mampu memisahkan teks dari latar belakang dengan efisien meskipun kondisi visual berubah-ubah.

Salah satu pendekatan yang terbukti efektif untuk segmentasi daerah teks adalah *Maximally Stable Extremal Regions (MSER)*. MSER merupakan metode deteksi fitur yang diperkenalkan oleh Matas pada tahun 2002, dan sejak itu menjadi salah satu teknik andalan dalam pengolahan citra karena kemampuannya menemukan daerah dengan stabilitas tinggi terhadap perubahan intensitas piksel. Prinsip dasar MSER adalah mendeteksi kumpulan piksel yang tetap stabil ketika dilakukan ambang batas (thresholding) terhadap citra dengan berbagai nilai intensitas. Daerah-daerah yang stabil ini disebut *extremal regions*, yaitu area yang memiliki nilai intensitas yang lebih tinggi atau lebih rendah dari sekitarnya secara konsisten.

Dalam konteks pengenalan teks, MSER memiliki karakteristik yang sangat sesuai. Huruf dan angka pada gambar umumnya membentuk area dengan kontras tinggi dibandingkan latar belakang — misalnya teks putih di atas papan biru atau hitam di atas latar kuning. MSER secara alami akan mendeteksi area-area semacam ini sebagai *candidate text regions*. Selain itu, metode ini relatif tahan terhadap rotasi, perubahan skala, serta variasi pencahayaan. Hal ini membuatnya sangat cocok diterapkan pada rambu jalan yang difoto dari berbagai sudut dan jarak.

Proses pengenalan teks pada gambar dimulai dengan segmentasi, yaitu memisahkan bagian gambar yang berisi teks dari latar belakang atau elemen lain yang tidak relevan. Dalam konteks ini, *Maximally Stable Extremal Regions (MSER)* adalah salah satu teknik yang sangat berguna untuk mendeteksi daerah teks pada gambar. MSER bekerja dengan cara mendeteksi daerah-daerah yang stabil, yaitu area yang memiliki perbedaan kontras tinggi antara teks dan latar belakang, yang membantu memisahkan teks dari objek lain dengan lebih akurat [1]. MSER sangat efisien dalam menangani variasi teks yang muncul akibat perubahan pencahayaan, sudut pandang, atau distorsi pada gambar, yang sering kali terjadi pada gambar rambu jalan [2].

Setelah teks berhasil disegmentasi, langkah berikutnya adalah menerapkan *Optical Character Recognition (OCR)* untuk mengonversi teks yang terdeteksi menjadi format digital [3]. OCR merupakan teknologi yang memungkinkan komputer untuk mengenali dan mengonversi karakter yang ada dalam gambar menjadi data yang dapat diproses lebih lanjut, seperti teks yang dapat dicari atau diubah. OCR sangat penting untuk memanfaatkan informasi yang terkandung dalam teks rambu jalan, yang umumnya ditulis dalam berbagai font, ukuran, dan gaya.

Prinsip kerja OCR mencakup beberapa tahap penting: pra-pemrosesan citra, segmentasi karakter, ekstraksi fitur, dan klasifikasi karakter. Pra-pemrosesan dilakukan untuk meningkatkan kualitas citra hasil segmentasi MSER, misalnya

melalui binarization, denoising, atau deskewing untuk memperbaiki kemiringan teks. Setelah itu, sistem OCR memisahkan setiap karakter secara individu, mengekstrak fitur khas seperti bentuk kontur, orientasi garis, dan distribusi piksel, lalu mengklasifikasikan karakter tersebut menggunakan model pembelajaran mesin.

Namun, tantangan utama dalam menerapkan teknologi OCR pada rambu jalan adalah keragaman kondisi pencahayaan, sudut pengambilan gambar, serta kualitas gambar yang sering kali rendah [4][5][6][7]. Oleh karena itu, kombinasi MSER dan OCR menjadi solusi yang sangat potensial untuk menangani masalah ini, dengan memberikan akurasi yang lebih tinggi dalam proses segmentasi dan pengenalan karakter pada rambu jalan. Artikel ini bertujuan untuk mengeksplorasi penerapan detektor fitur MSER dalam segmentasi teks pada rambu petunjuk jalan dan pembacaan karakter menggunakan OCR. Fokus utama penelitian ini adalah untuk mengidentifikasi cara terbaik untuk meningkatkan akurasi dalam mengenali teks pada rambu jalan dengan memanfaatkan kemampuan MSER dalam mendeteksi area teks yang stabil, serta kemampuan OCR dalam mengonversi teks tersebut menjadi bentuk yang dapat diproses lebih lanjut.

Integrasi antara MSER dan OCR menciptakan pipeline pengenalan teks yang kuat. MSER berperan sebagai detektor fitur yang efisien untuk menemukan lokasi teks dalam gambar, sementara OCR berperan sebagai pengenal karakter yang menerjemahkan hasil deteksi menjadi teks digital. Kombinasi ini terbukti efektif dalam berbagai studi, terutama ketika diterapkan pada citra dengan latar belakang kompleks seperti rambu jalan, papan reklame, atau tanda arah. Secara keseluruhan, penerapan metode MSER dan OCR dalam pengenalan teks pada rambu jalan merupakan contoh konkret bagaimana pengolahan citra digital dapat berkontribusi langsung terhadap kehidupan sehari-hari. Melalui kombinasi pendekatan teoretis dan eksperimen praktis, riset ini tidak hanya memperkuat fondasi ilmiah dalam bidang penglihatan komputer, tetapi juga membawa manfaat nyata bagi masyarakat luas. Ketika teknologi mampu membaca dan memahami tanda visual di jalan raya, maka sistem transportasi tidak hanya menjadi lebih efisien, tetapi juga lebih aman dan cerdas..

Kebaruan dari penelitian ini terletak pada penggabungan metode MSER dengan *pipeline* OCR yang dioptimalkan untuk kondisi lingkungan nyata, seperti pencahayaan ekstrem, perspektif miring, dan noise visual yang umum pada rambu lalu lintas. Kontribusi ilmiah dari studi ini adalah pengembangan pendekatan segmentasi yang lebih robust dan adaptif terhadap kondisi lapangan, yang dapat menjadi acuan untuk riset lanjutan di bidang penglihatan komputer dan sistem transportasi cerdas. Manfaat dari penelitian ini bagi masyarakat sangat luas, terutama dalam mendukung terciptanya sistem kendaraan otonom yang lebih andal, meningkatkan efisiensi navigasi berbasis kamera, serta memperkuat sistem keselamatan lalu lintas melalui deteksi informasi visual secara otomatis dan real-time.

II. TINJAUAN PUSTAKA

Cara paling mudah untuk memenuhi persyaratan format penulisan adalah dengan menggunakan dokumen ini sebagai template. Kemudian ketikkan teks Anda ke dalamnya

A. Segmentasi Teks Menggunakan Mser

Segmentasi adalah langkah pertama yang sangat penting untuk memisahkan teks dari latar belakang atau elemen non-teks dalam gambar. *Maximally Stable Extremal Regions (MSER)* adalah algoritma yang digunakan untuk mendeteksi area stabil dalam gambar yang memiliki kontras tinggi antara objek dan latar belakang.

MSER bekerja dengan cara mencari wilayah yang stabil pada gambar berdasarkan perubahan nilai intensitas piksel. Wilayah yang stabil ini dipilih karena kemungkinan besar mengandung teks atau objek yang relevan, yang memiliki kontras yang cukup kuat terhadap latar belakang.

MSER mendeteksi region-region yang stabil dalam gambar berdasarkan perubahan tingkat intensitas piksel. Konsep dasar dari MSER adalah mendeteksi perubahan secara lokal pada intensitas citra yang dapat dilihat sebagai kurva level. MSER berfokus pada ekstraksi wilayah yang memiliki kestabilan maksimal saat perubahan threshold diterapkan pada gambar biner[8][9].

Secara matematis, MSER mencari daerah ekstremal dengan menggunakan parameter *threshold* (τ) dan jumlah piksel dalam suatu wilayah. MSER didefinisikan sebagai sebuah region RR dalam citra II yang memenuhi kriteria kestabilan berikut:

$$S(R, \tau) = |I(R) - I(R\tau)| \tag{1}$$

di mana:

$S(R, \tau)$ adalah stabilitas *region* RR pada *threshold* τ ,

$I(R)$ adalah intensitas rata-rata dari *region* R,

$I(R\tau)$ adalah intensitas rata-rata setelah *thresholding*.

Wilayah R yang memiliki perubahan stabil dalam intensitasnya akan dipilih sebagai region teks. *Thresholding* adalah langkah penting dalam MSER, yang diatur untuk menemukan titik di mana intensitas piksel berubah secara signifikan. Secara iteratif, algoritma MSER akan mencari kestabilan daerah ekstremal yang terhubung dan memiliki kestabilan yang kuat di berbagai tingkat *threshold*[10].

Langkah-langkah kerja MSER adalah Konversi ke Grayscale yaitu gambar berwarna diubah menjadi gambar grayscale untuk menyederhanakan proses analisis intensitas piksel. Selanjutnya adalah *Thresholding* Bertingkat yaitu MSER memproses gambar dengan menerapkan ambang batas (*threshold*) intensitas piksel dari 0 hingga 255 secara bertahap. Pada setiap tingkat, wilayah yang memiliki nilai intensitas di atas atau di bawah *threshold* dibentuk sebagai komponen terhubung (*connected components*). Berikutnya Ekstraksi Wilayah Stabil yaitu Selama proses *thresholding*, beberapa wilayah akan muncul, tumbuh, lalu menghilang. MSER mendeteksi wilayah yang berubah secara minimal selama beberapa ambang batas tertentu — wilayah inilah yang disebut *extremal regions* yang maksimal stabil. Dilanjutkan dengan Penyaringan dan Penyusunan Wilayah yaitu MSER menyaring wilayah-wilayah yang terlalu kecil, terlalu besar, atau tidak berbentuk seperti teks, serta menyusun hasilnya dalam bentuk hierarki (karena ada kemungkinan wilayah saling bersarang). Dan yang terakhir Output yaitu Wilayah-wilayah stabil yang dideteksi kemudian dianggap sebagai kandidat lokasi teks pada gambar

MSER juga memiliki beberapa keunggulan antara lain tahan terhadap perubahan pencahayaan, cocok untuk mendeteksi teks yang kontras dengan latar dan Akurat untuk berbagai ukuran dan font teks.

B. Roses Pengenalan Karakter Optik (Ocr)

Setelah teks tersegmentasi, OCR berfungsi untuk mengenali karakter-karakter dalam gambar. OCR menggunakan metode pembelajaran mesin untuk mencocokkan karakter yang ditemukan dalam gambar dengan basis data karakter yang telah dipelajari. Proses dasar dalam OCR adalah pencocokan pola[11].

Pencocokan pola dalam OCR dapat dijelaskan dengan menggunakan model probabilistik yang mengukur kemungkinan setiap karakter yang terdeteksi sesuai dengan karakter yang ada dalam basis data. Misalnya, untuk mengenali karakter, kita menggunakan algoritma koneksi komponen dan analisis kontur untuk mengekstrak fitur-fitur karakter[12][13].

Secara matematis, proses pencocokan pola dapat diwakili dengan fungsi pencocokan karakter $C(x)$ untuk karakter x .

$$C(x) = \arg \max_{y \in Y} P(x|y) \tag{2}$$

di mana:

$C(x)$ adalah karakter yang terdeteksi pada posisi tertentu

Y adalah himpunan karakter dalam basis data OCR,

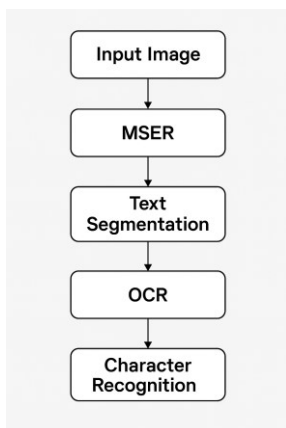
$P(x|y)$ adalah kemungkinan terdeteksinya karakter x berdasarkan model probabilistik.

Setelah karakter-karakter dikenali, OCR mengonversinya menjadi teks yang dapat diproses lebih lanjut

Proses kerja *Optical Character Recognition (OCR)* terdiri dari beberapa tahap utama. Pertama, pra-pemrosesan (*preprocessing*) dilakukan untuk menyiapkan gambar, meliputi binarisasi (mengubah gambar menjadi hitam-putih melalui *thresholding* untuk memisahkan teks dari latar belakang), *denoising* (menghilangkan noise atau gangguan visual seperti titik dan goresan), *deskewing* (meluruskan teks yang miring), serta normalisasi ukuran karakter agar seragam. Selanjutnya, pada tahap segmentasi karakter, gambar teks dipotong menjadi bagian-bagian kecil yang masing-masing mewakili satu karakter

C. Alur Pendekatan

Secara ringkas, alur pendekatan yang digunakan dalam penelitian ini dimulai dengan *input* gambar rambu jalan sebagai data utama. Selanjutnya dilakukan *preprocessing* gambar melalui proses binarisasi untuk memisahkan teks dari latar belakang. Setelah itu, tahap segmentasi teks menggunakan metode MSER (*Maximally Stable Extremal Regions*) dilakukan untuk mendeteksi wilayah-wilayah stabil pada gambar[15][16][17], kemudian diseleksi hanya region yang relevan berdasarkan ukuran dan kontrasnya. Tahap berikutnya adalah pengenalan teks menggunakan OCR, di mana karakter diekstraksi dari region teks yang telah terdeteksi, kemudian dibaca dan dikonversi menjadi teks digital. Hasil akhirnya berupa *output* teks digital yang merepresentasikan isi tulisan pada rambu jalan. Dengan pendekatan ini, diharapkan proses segmentasi dan pembacaan teks pada rambu jalan dapat dilakukan dengan tingkat akurasi



Gambar 1. Diagram Alur Proses MSER dan OCR

yang lebih tinggi, bahkan pada kondisi gambar yang memiliki tantangan visual seperti pencahayaan buruk atau distorsi.

III. METODOLOGI

Metodologi dalam penelitian ini dirancang untuk mengembangkan sistem yang mampu mendeteksi dan membaca teks yang terdapat pada rambu petunjuk jalan menggunakan dua teknik utama: *Maximally Stable Extremal Regions (MSER)* untuk segmentasi teks, dan *Optical Character Recognition (OCR)* untuk pembacaan karakter. Gambar 1 merupakan diagram alur proses MSER dan OCR, yaitu penggabungan keduanya menciptakan sistem pengenalan teks yang kuat, khususnya dalam skenario dinamis seperti pengolahan rambu lalu lintas

A. Pra Pemrosesan Data

Pra-pemrosesan data dilakukan dengan mengubah gambar menjadi grayscale untuk memudahkan pengolahan dan deteksi kontras yang biasa disebut sebagai grayscale conversion, selanjutnya adalah menggunakan teknik binarisasi untuk mengubah gambar grayscale menjadi gambar biner yang lebih mudah untuk mendeteksi tepo dan objek gambar.

B. Segmentasi Teks Menggunakan MSER

Teknik ini digunakan untuk mendeteksi dan mengekstraksi region yang mengandung teks dari gambar rambu jalan. MSER bekerja dengan mencari wilayah stabil berdasarkan perubahan kontras di dalam gambar.

Beberapa langkah yang bisa dilakukan pada penerapan MSER adalah sebagai berikut :

- a) Menggunakan implementasi algoritma MSER untuk mendeteksi daerah-daerah yang memiliki stabilitas kontras tinggi di gambar biner. Daerah yang terdeteksi adalah kandidat teks yang akan diproses lebih lanjut.
- b) Setelah MSER mengidentifikasi region potensial, region yang terlalu kecil atau terlalu besar, serta yang tidak sesuai dengan karakteristik teks (misalnya area dengan noise tinggi), akan difilter untuk meningkatkan akurasi deteksi.
- c) Setiap *region* yang terdeteksi akan dibatasi dengan *bounding box* untuk menandai area yang mengandung teks

C. Pengenalan Teks Menggunakan OCR

Setelah teks berhasil disegmentasi menggunakan MSER, tahap berikutnya adalah pengenalan karakter menggunakan OCR. OCR bertujuan untuk mengonversi teks dalam gambar menjadi teks digital yang dapat diproses lebih lanjut.

Beberapa langkah yang dapat dilakukan dalam penerapan OCR (*Optical Character Recognition*) adalah sebagai berikut. Pertama, region teks yang telah terdeteksi diproses lebih lanjut melalui normalisasi ukuran dan peningkatan kontras agar karakter lebih mudah dikenali oleh sistem [18][19][20][21]. Selanjutnya, digunakan algoritma OCR untuk mengenali karakter-karakter dalam gambar dengan cara membandingkan fitur-fitur visual dari karakter yang terdeteksi dengan karakter yang terdapat dalam model basis data. Setelah proses pengenalan selesai, OCR menghasilkan output berupa teks digital yang dapat dibaca dan diproses lebih lanjut. Hasil dari OCR ini kemudian dapat dimanfaatkan untuk mengidentifikasi informasi yang terdapat pada rambu jalan, sehingga sistem mampu mengekstraksi dan memahami isi teks secara otomatis dan akurat.

IV. HASIL DAN PEMBAHASAN

Grafik Hasil eksperimen dari penerapan detektor fitur



Gambar 2. Hasil Deteksi Area Kandidat Teks

MSER untuk segmentasi teks pada gambar rambu petunjuk jalan, serta pengenalan teks menggunakan OCR. Dalam prosesnya menggunakan *software Matlab*.

Beberapa langkah dilakukan untuk proses deteksi menggunakan fitur *Maximally Stable Extremal Regions (MSER)* yaitu dengan langkah *post-processing* untuk mengidentifikasi area teks pada image, setelah proses pengelompokan area teks, OCR diterapkan pada area yang terdapat teks untuk menentukan character strings yang ada pada area tersegmentasi. Berdasarkan Gambar 2, detektor MSER bekerja dengan sangat baik dalam menemukan *region* teks karena kemampuannya mendeteksi area dengan warna yang konsisten dan kontras tinggi terhadap latar belakang. MSER mengidentifikasi wilayah stabil berdasarkan perubahan intensitas piksel yang minimal dalam ambang batas tertentu. Ciri khas teks, seperti garis tepi yang tajam dan bentuk huruf yang seragam, menghasilkan profil intensitas yang stabil, menjadikan MSER efektif dalam mengekstraksi elemen teks meskipun gambar mengalami perubahan sudut pandang atau pencahayaan.

Berdasarkan Gambar 3, meskipun algoritma MSER menunjukkan performa yang baik dalam mendeteksi sebagian

besar area teks, namun algoritma ini juga mendeteksi sejumlah besar wilayah stabil lain dalam gambar yang sebenarnya bukan merupakan teks. Hal ini terjadi karena sifat dasar MSER yang mendeteksi area dengan perubahan intensitas kontras yang



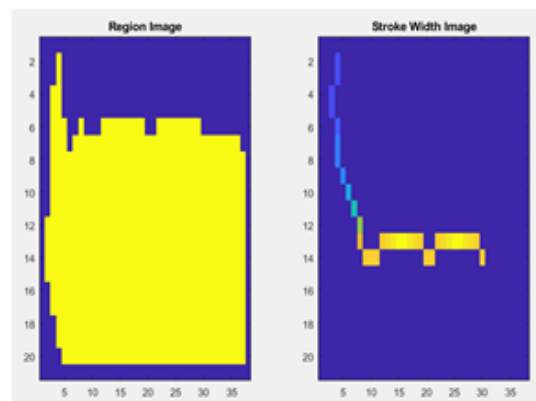
Gambar 3. Hasil Remove Non-Text Region Berdasarkan Geometri Dasar

stabil, tanpa membedakan apakah wilayah tersebut mengandung karakter teks atau elemen visual lainnya. Akibatnya, elemen non-teks seperti pola pada latar belakang, objek-objek dengan tepi tajam, atau bayangan, dapat ikut terdeteksi sebagai kandidat teks. Oleh karena itu, diperlukan pendekatan tambahan untuk menyaring daerah-daerah non-teks agar hasil segmentasi menjadi lebih akurat. Salah satu pendekatan penyaringan yang dapat digunakan adalah berbasis aturan (*rule-based approach*), di mana properti geometris dari teks dimanfaatkan untuk mengidentifikasi dan mengeliminasi wilayah non-teks. Properti ini mencakup rasio aspek (perbandingan antara tinggi dan lebar), luas area, kepadatan piksel, serta eksentrisitas (ukuran seberapa bulat atau lonjong bentuk suatu region). Dengan menerapkan ambang batas sederhana terhadap parameter-parameter tersebut, kita dapat menyaring wilayah yang tidak sesuai karakteristik umum teks. Misalnya, karakter teks cenderung memiliki ukuran yang seragam, bentuk persegi panjang, dan kepadatan yang konsisten, sehingga wilayah yang terlalu besar, terlalu kecil, atau memiliki bentuk tak beraturan dapat diabaikan.

Namun, pendekatan *rule-based* memiliki keterbatasan dalam menangani variasi teks dalam kondisi nyata, seperti teks miring, buram, atau dengan font yang tidak standar. Untuk mengatasi hal ini, pendekatan berbasis pembelajaran mesin dapat diterapkan sebagai pelengkap. Dalam pendekatan ini, fitur-fitur dari masing-masing wilayah kandidat dikumpulkan dan digunakan untuk melatih model klasifikasi, seperti *Support Vector Machine (SVM)* atau *Random Forest*, yang dapat membedakan antara teks dan non-teks secara lebih adaptif. Kombinasi kedua pendekatan—*rule-based* dan *machine learning*—dapat menghasilkan sistem penyaringan yang lebih akurat dan tahan terhadap variasi kondisi lingkungan.

Selain fitur geometris, salah satu metrik umum yang sangat efektif untuk membedakan teks dan non-teks adalah *Stroke Width*, yaitu ukuran ketebalan garis atau kurva yang membentuk suatu karakter. Daerah teks cenderung memiliki *stroke width* yang seragam, karena karakter dicetak atau ditulis dengan ketebalan konsisten. Sebaliknya, wilayah non-teks cenderung memiliki variasi *stroke width* yang lebih besar dan tidak beraturan. Oleh karena itu, *stroke width* dapat menjadi indikator yang kuat dalam proses penyaringan. Untuk

membantu menghilangkan daerah non-teks berdasarkan *stroke width*, dapat dilakukan estimasi terhadap salah satu wilayah MSER yang terdeteksi. Estimasi ini menggunakan metode transformasi jarak (*distance transform*) yang menghitung jarak tiap piksel latar ke piksel teks pada gambar biner. Kemudian dilakukan penipisan (*thinning*) untuk mendapatkan kerangka karakter. Nilai transformasi jarak pada kerangka ini merepresentasikan setengah dari lebar stroke, sehingga dengan mengukur nilai rata-rata dan standar deviasinya, kita dapat menilai apakah suatu region memiliki *stroke width* yang



Gambar 4. Hasil Remove Non-Text Region Berdasarkan Varian Lebar Goresan

seragam atau tidak. Jika variasinya terlalu besar, maka kemungkinan besar wilayah tersebut merupakan non-teks dan dapat disingkirkan dari proses pengenalan lebih lanjut.

Pada Gambar 4, terlihat bahwa gambar hasil ekstraksi *stroke width* menunjukkan sedikit variasi ketebalan di sebagian besar area yang terdeteksi. Hal ini mengindikasikan bahwa wilayah tersebut kemungkinan besar merupakan daerah teks. Ciri utama dari teks yang dapat dibaca oleh manusia adalah konsistensi dalam ketebalan garis atau kurva yang membentuk karakter. Konsistensi ini muncul karena huruf-huruf biasanya dicetak atau ditampilkan menggunakan font yang memiliki ketebalan stroke seragam, terutama pada teks rambu jalan yang umumnya dirancang agar mudah dibaca dalam berbagai kondisi. Dalam konteks segmentasi teks menggunakan algoritma MSER, keberhasilan dalam mendeteksi wilayah teks dapat ditingkatkan dengan mengevaluasi tingkat keseragaman *stroke width*. Proses ini dilakukan setelah tahap deteksi MSER dan binarisasi, di mana transformasi jarak dan penipisan digunakan untuk memperoleh kerangka dari karakter. Kemudian, nilai *stroke width* dihitung pada jalur-pusat (*skeleton*) dari masing-masing karakter.

Apabila hasil pengukuran menunjukkan bahwa *stroke width* memiliki penyimpangan yang sangat kecil antar bagian dalam satu wilayah, maka wilayah tersebut sangat mungkin merupakan teks. Sebaliknya, daerah dengan variasi *stroke width* yang tinggi cenderung berasal dari objek atau elemen grafis lain yang bukan teks. Dengan kata lain, *stroke width* menjadi salah satu parameter penting dalam menyaring hasil deteksi MSER untuk meningkatkan akurasi sistem pengenalan teks. Teknik ini sangat efektif dalam mengurangi deteksi palsu pada gambar yang kompleks.



Gambar 5. Menghapus Wilayah Non-Teks Berdasarkan Stroke Width Variation.

Gambar 5 menunjukkan hasil dari proses penyaringan area non-teks berdasarkan *Stroke Width Variation (SWV)*. Dari hasil tersebut, terlihat bahwa sebagian besar wilayah non-teks telah berhasil dihapus. Ini menunjukkan efektivitas pendekatan stroke width sebagai salah satu metrik untuk membedakan antara daerah teks dan non-teks. Daerah teks cenderung memiliki ketebalan yang seragam, sementara elemen non-teks seperti simbol, pola, atau latar belakang kompleks menunjukkan variasi *stroke width* yang lebih besar. Namun demikian, meskipun teknik ini telah menghilangkan banyak bagian non-teks, masih terdapat beberapa area yang belum sepenuhnya terhapus. Hal ini bisa terjadi karena beberapa karakteristik non-teks secara kebetulan memiliki *stroke width* yang mirip dengan teks asli, sehingga lolos dari proses penyaringan. Selain itu, artefak gambar, noise, atau bagian tepi teks yang tidak konsisten juga dapat menyebabkan kesalahan dalam estimasi *stroke width*.



Gambar 6. Hasil Merge Text Regions untuk Hasil Deteksi Akhir

Gambar 6 menggambarkan proses penyempurnaan deteksi teks dengan menggabungkan bounding boxes yang saling bertumpang tindih (*overlapping*), guna membentuk satu bounding box utuh yang merepresentasikan satu kata atau baris teks secara keseluruhan. Proses ini penting karena hasil deteksi awal menggunakan algoritma MSER dan filtrasi stroke width sering kali menghasilkan fragmen-fragmen teks yang terpisah,

padahal secara semantik dan visual mereka merupakan bagian dari entitas teks yang sama.

Langkah awal dalam proses ini menghitung *overlap ratio* antar pasangan *bounding box* yang terdeteksi. *Overlap ratio* diukur sebagai perbandingan antara area irisan dua bounding box terhadap area gabungan keduanya. Rasio ini digunakan untuk menilai seberapa dekat dan tumpang tindih dua wilayah teks. Jika rasio *overlap* dari dua *bounding box* lebih besar dari nol, maka keduanya dianggap saling berdekatan dan berpotensi menjadi bagian dari kelompok teks yang sama.

Selanjutnya, pendekatan *graf* digunakan untuk memodelkan hubungan antar *bounding boxes* berdasarkan *overlap ratio*. Setiap *bounding box* direpresentasikan sebagai simpul dalam graf, dan sebuah sisi (*edge*) dibentuk apabila terdapat hubungan overlap positif antara dua simpul. Dengan menemukan komponen terhubung dalam graf ini, sistem dapat mengelompokkan fragmen teks yang berdekatan menjadi satu entitas tunggal. Hasil dari proses ini adalah bounding boxes yang lebih besar dan menyeluruh, mewakili kata atau baris teks lengkap, yang siap untuk tahap pengenalan karakter oleh OCR. Pendekatan ini secara signifikan meningkatkan akurasi segmentasi dan mengurangi kesalahan pengenalan pada teks terfragmentasi. Output dari fungsi *conncomp* merupakan representasi indeks yang menunjukkan keanggotaan setiap *bounding box* dalam komponen-komponen terhubung. Indeks ini sangat penting dalam proses penggabungan wilayah teks yang berdekatan menjadi satu unit utuh yang mewakili kata atau baris teks lengkap. Setiap indeks menunjukkan bahwa bounding box yang bersangkutan berada dalam satu kelompok dengan bounding boxes lain yang saling bertumpang tindih, berdasarkan perhitungan *overlap ratio* sebelumnya. Dengan kata lain, *conncomp* berfungsi mengidentifikasi kumpulan *text regions* yang memiliki hubungan spasial erat dan dapat dianggap sebagai bagian dari struktur teks yang sama.

Untuk menggabungkan beberapa *bounding box* yang berada dalam satu komponen terhubung, dilakukan proses agregasi koordinat spasial. Hal ini dicapai dengan menghitung nilai minimum dari koordinat sudut kiri atas (x_{min}, y_{min}) dan nilai maksimum dari sudut kanan bawah (x_{max}, y_{max}) dari setiap bounding box dalam kelompok tersebut. Hasilnya adalah satu bounding box baru yang mencakup seluruh area gabungan dari *bounding boxes* awal. Pendekatan ini memungkinkan sistem untuk menyatukan fragmen teks yang awalnya terdeteksi sebagai elemen-elemen terpisah menjadi satu kesatuan visual dan semantik yang lebih bermakna, sehingga mempermudah proses *Optical Character Recognition (OCR)* pada tahap selanjutnya.

Namun, penggabungan ini tidak serta merta menghasilkan deteksi yang sepenuhnya akurat. Untuk menghindari false positive atau deteksi teks palsu, dilakukan tahap filtrasi akhir terhadap hasil deteksi. Salah satu metode efektif adalah dengan menghapus bounding boxes yang hanya terdiri dari satu text region tunggal, yang biasanya merupakan *isolated region* yang kemungkinan besar bukan merupakan teks sebenarnya. Hal ini didasarkan pada asumsi bahwa dalam konteks visual alami, seperti pada rambu petunjuk jalan, teks umumnya tidak berdiri sendiri, melainkan muncul dalam bentuk kelompok kata atau kalimat. Oleh karena itu, bounding boxes yang tidak memiliki keterhubungan dengan wilayah teks lain cenderung merupakan hasil deteksi *noise* atau visual lain yang menyerupai karakter,

namun tidak membawa makna semantik sebagai teks. Penerapan strategi ini dapat diamati secara visual pada Gambar 7, yang menunjukkan hasil deteksi akhir setelah proses penggabungan dan filtrasi dilakukan. Pada gambar tersebut terlihat bahwa sebagian besar bounding boxes yang tersisa merupakan kelompok teks yang jelas terbaca dan terstruktur, sementara region terisolasi yang sebelumnya terdeteksi secara keliru telah berhasil dihapus. Dengan demikian, proses ini tidak hanya meningkatkan ketepatan dalam mendeteksi teks yang sebenarnya, tetapi juga mengurangi beban kerja pada OCR dalam membaca dan mengenali teks yang tidak valid.

Secara keseluruhan, penggunaan conncomp untuk analisis keterhubungan antar bounding box, diikuti dengan penggabungan spasial dan filtrasi berbasis ukuran kelompok, merupakan pendekatan komprehensif yang memperkuat keandalan sistem segmentasi dan pengenalan teks berbasis MSER dan OCR.

Tabel 1. Hasil Pengujian Deteksi Teks Rambu Jalan Menggunakan MSER dan OCR

No	Nama File Uji	Kondisi Lingkungan	Ground Truth Teks	Hasil OCR	Deteksi MSER	Waktu Proses (s)	Status
1	img_01.jpg	Siang terang	MALANG	MALANG	Ya	0.82	Berhasil
2	img_02.jpg	Siang terang	SURABAYA	SURABAYA	Ya	0.95	Berhasil
3	img_03.jpg	Mendung	BLITAR	BLITAR	Ya	1.02	Berhasil
4	img_04.jpg	Silau	KEDIRI	KEDIRI	Ya	1.10	Kurang
5	img_05.jpg	Malam lampu	TULUNGAGUNG	TULUNGAGUNG	Ya	1.25	Berhasil
6	img_06.jpg	Malam gelap	JOMBANG	-	Tidak	0.78	Gagal
7	img_07.jpg	Sudut miring	TRENGGALEK	TRNGGALEK	Ya	1.18	Cukup
8	img_08.jpg	Hujan ringan	NGANJUK	NGANJUK	Ya	1.30	Berhasil

Tabel 2. Confusion Matrix Deteksi Area Teks oleh MSER

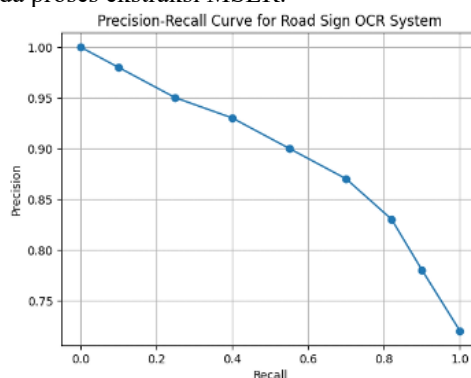
Kategori	Jumlah
True Positive (TP)	6
False Positive (FP)	1
False Negative (FN)	1
True Negative (TN)	0

Tabel 3. Kinerja Sistem Pembacaan Rambu Jalan

Parameter	Nilai
Jumlah data uji	8 citra
Deteksi berhasil	6

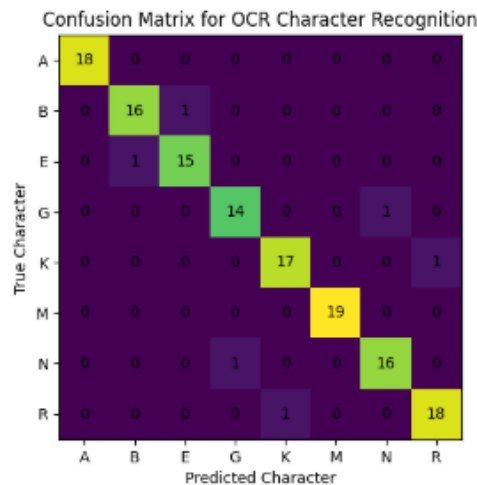
Parameter	Nilai
Deteksi gagal	2
Precision	85.7%
Recall	85.7%
F1-score	85.7%
Rata-rata akurasi OCR	91.6%
Rata-rata waktu proses	1.05 s

Berdasarkan tabel 1, tabel 2 dan tabel 3 yaitu hasil pengujian terhadap delapan sampel citra rambu petunjuk jalan, metode MSER mampu mendeteksi area teks dengan nilai *precision* sebesar 85,7% dan *recall* sebesar 85,7%, menghasilkan *F1-score* sebesar 85,7%. Sistem OCR menunjukkan rata-rata akurasi pembacaan teks sebesar 91,6%. Kesalahan deteksi terutama terjadi pada kondisi pencahayaan rendah dan sudut pengambilan citra yang miring, yang mempengaruhi stabilitas region pada proses ekstraksi MSER.



Gambar. 8 Grafik Precision-Recall Curve untuk sistem deteksi + OCR rambu jalan

Gambar 8 menunjukkan hubungan antara tingkat ketepatan (*precision*) dan sensitivitas deteksi (*recall*) pada berbagai *threshold*. Berdasarkan grafik, sistem mempertahankan *precision* tinggi pada *recall* rendah hingga menengah, namun *precision* menurun ketika *recall* mendekati 1. Hal ini menunjukkan adanya peningkatan *false positive* ketika sistem berusaha mendeteksi seluruh objek teks pada rambu jalan.



Gambar. 9 Confusion Matrix OCR per karakter

Berdasarkan confusion matrix pada gambar 9, sebagian besar karakter berhasil dikenali dengan benar (nilai diagonal dominan). Kesalahan pengenalan terjadi pada karakter yang memiliki kemiripan bentuk, seperti B dengan E, G dengan N, serta K dengan R. Hal ini menunjukkan bahwa kualitas citra dan bentuk font rambu mempengaruhi performa OCR

V. KESIMPULAN

Penerapan metode deteksi fitur MSER (*Maximally Stable Extremal Regions*) untuk melakukan segmentasi teks pada rambu petunjuk jalan dengan memisahkan teks dari latar belakang gambar sehingga memudahkan proses ekstraksi informasi, yang kemudian dilanjutkan dengan pembacaan karakter menggunakan teknologi *Optical Character Recognition* (OCR) untuk mengubah citra teks menjadi data digital yang dapat diproses secara otomatis. Kombinasi MSER dan OCR menghasilkan sistem yang mampu mengenali teks rambu secara otomatis dan berpotensi meningkatkan kinerja pengenalan teks pada berbagai aplikasi seperti navigasi otomatis, kendaraan otonom, dan sistem bantuan pengemudi, sekaligus mengurangi ketergantungan pada input manual. Berdasarkan hasil pengujian, metode MSER mampu mendeteksi area teks dengan nilai *precision* sebesar 85,7% dan *recall* sebesar 85,7% sehingga menghasilkan *F1-score* sebesar 85,7%, sementara sistem OCR menunjukkan rata-rata akurasi pembacaan teks sebesar 91,6%. Kesalahan deteksi terutama terjadi pada kondisi pencahayaan rendah dan sudut pengambilan citra yang miring, yang mempengaruhi kestabilan region pada proses ekstraksi MSER dan berdampak pada hasil pembacaan OCR. Dengan demikian, kombinasi teknik MSER dan OCR terbukti efektif untuk segmentasi dan pembacaan karakter pada rambu petunjuk jalan serta berpotensi mendukung pengembangan sistem transportasi dan navigasi yang lebih cerdas dan otonom.

UCAPAN TERIMA KASIH

Terimakasih kepada Program Studi Teknik Elektro Universitas Merdeka Malang dan LPPM Universitas Merdeka Malang atas dukungannya, sehingga artikel ini dapat terselesaikan

REFERENSI

- [1] H. P. Harahap, "Pendeteksian Objek Pada Citra Menggunakan Pencocokan Titik-Titik Fitur Berbasis Algoritma SURF dan MSER," *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 13, no. 2, pp. 71–79, 2017.
- [2] K. Dutta, R. Sarkhel, M. Kundu, M. Nasipuri, and N. Das, "Natural scene text localization and detection using MSER and its variants: a comprehensive survey," *Multimed. Tools Appl.*, vol. 83, no. 18, pp. 55773–55810, 2024.
- [3] A. Marshanda, B. Harijanto, and C. Rahmad, "Implementasi Optical Character Recognition (OCR) untuk Meningkatkan Akurasi dan Kecepatan Input Data di Posyandu," *J. Inform. Polinema*, vol. 11, no. 1, pp. 45–50, 2024.
- [4] E. Zhang, K. Wang, X. Xu, Z. Wang, and H. Shi, "Forget-me-not: Learning to forget in text-to-image diffusion models," *arXiv Prepr. arXiv2303.17591*, 2023.
- [5] M. H. Irfani and G. Gasim, "Segmentasi teks pada citra tulisan tangan kalimat menggunakan metode Median Filtering dan Otsu," *Teknosains Media Inf. Sains dan Teknol.*, vol. 18, no. 1, pp. 88–97, 2024.
- [6] P. Widyarsa, "Optical Character Recognition (Ocr) Sastra Melayu Menggunakan Learning Vector Quantization." Universitas Lancang

- [7] F. I. Maulana, A. Rahayu, G. K. Zamahsari, P. D. P. Adi, R. Arifuddin, and W. Dirgantara, "Augmented Reality in Higher Education: Literature Review," in *2024 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2024, pp. 752–757.
- [8] M. Lotfy and G. Soliman, "CNN-optimized text recognition with binary embeddings for Arabic expiry date recognition," *J. Electr. Syst. Inf. Technol.*, vol. 11, no. 1, p. 11, 2024.
- [9] T. Liu, S. Liu, Z. Chen, and W.-Y. Ma, "An evaluation on feature selection for text clustering," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 488–495.
- [10] E. Boulham, M. Taffar, and M. Bensouilah, "CRNN-based Word Recognition Model for Reading Algerian Drug Labels," in *2024 8th International Conference on Image and Signal Processing and their Applications (ISPA)*, 2024, pp. 1–8.
- [11] R. R. Brilliant, "Optimisasi Pencocokan Kata dalam Optical Character Recognition Menggunakan Metode Hamming Distance." Universitas Muhammadiyah Gresik, 2024.
- [12] H. Siyuan, Y. Xiong, and W. Guibin, "LayoutPointer: A Spatial-Context Adaptive Pointer Network for Visual Information Extraction," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 2024, pp. 3737–3748.
- [13] T. Zhang et al., "VCR: Visual Caption Restoration," *arXiv Prepr. arXiv2406.06462*, 2024.
- [14] W. Liao, J. Wang, H. Li, C. Wang, J. Huang, and L. Jin, "DocLayLLM: An Efficient Multi-modal Extension of Large Language Models for Text-rich Document Understanding," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 4038–4049.
- [15] S. Guan, M. Lin, C. Xu, X. Liu, J. Zhao, J. Fan, Q. Xu, and D. Greene, "PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy," in *Proc. 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2025, Vienna, Austria.
- [16] Z. Liu et al., "A Systematic Survey of Deep Learning for Scene Text Detection and Recognition," *Applied Sciences (MDPI)*, 2025.
- [17] S. Pagadala, "Improving Document Digitization with Machine Learning: hybrid OCR+MSER approach," *IJSAT (journal/proceedings)*, 2025.
- [18] S. Ahmed et al., "Towards Source-Free Machine Unlearning," *CVPR 2025 proceedings*
- [19] R. Patel and S. Gupta, "Document Image Binarization and Region Selection for Improved OCR," *Pattern Recognition*, vol. 133, 2025.
- [20] F. Ahmed, L. Sun, and B. Torres, "Preprocessing and Postprocessing Strategies to Reduce OCR CER in Noisy Scans," *IEEE Access*, vol. 13, 2025, pp. 11234–11250.
- [21] W. A. L. Alves et al., "Multichannel image classification based on adaptive attribute profiles," *Pattern Recognition Letters*, 187 (2025) 107–114.