

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seleksi Nasional Berdasarkan Prestasi (SNBP) adalah jalur penerimaan mahasiswa baru yang memberikan kesempatan kepada siswa dengan prestasi akademik dan non-akademik untuk diterima di perguruan tinggi negeri tanpa melalui tes tertulis [1]. Dalam proses seleksi ini, sertifikat prestasi menjadi salah satu dokumen pendukung yang menunjukkan capaian siswa, baik berupa juara lomba, penghargaan akademik, maupun pencapaian di bidang lain yang relevan [2]. Secara teori, pengakuan terhadap prestasi yang dibuktikan melalui sertifikat merupakan bagian dari evaluasi holistik siswa, yaitu proses penilaian yang tidak hanya mempertimbangkan nilai akademik, tetapi juga capaian dan potensi siswa di luar aspek akademik [3]. Peningkatan jumlah peserta SNBP setiap tahun berdampak pada volume sertifikat yang harus diverifikasi oleh panitia seleksi [4]. Proses verifikasi manual ini memerlukan waktu yang panjang karena setiap sertifikat harus diperiksa satu per satu. Variasi desain, resolusi, serta kualitas hasil pemindaian atau citra sertifikat yang berbeda-beda turut memperlambat proses pemeriksaan dan meningkatkan potensi terjadinya kesalahan identifikasi dokumen [5].

Meskipun sertifikat prestasi memiliki peran penting, sistem verifikasi yang digunakan dalam SNBP saat ini belum optimal [6]. Data dari Panitia SNPMB mencatat bahwa pada tahun 2024 terdapat 702.312 peserta SNBP dan pada tahun 2026 tercatat 776.515 peserta SNBP, dengan setiap peserta dapat mengunggah lebih dari satu sertifikat prestasi [7], [8]. Sehingga, potensi dokumen yang harus diperiksa panitia mencapai lebih dari dua juta berkas dalam satu periode seleksi. Volume dokumen yang besar ini, ditambah variasi format dan desain sertifikat, menjadi tantangan besar bagi proses verifikasi manual [9]. Selain itu, terdapat kemungkinan ditemukannya sertifikat yang tidak kredibel berdasarkan parameter visual yang digunakan dalam proses verifikasi. Oleh sebab itu, dibutuhkan pendekatan berbasis teknologi yang mampu mengotomatisasi proses identifikasi parameter visual pada

sertifikat prestasi sehingga proses identifikasi sertifikat dapat dilakukan secara cepat, akurat, dan konsisten. Penelitian ini bertujuan untuk mengembangkan sistem yang mampu mendeteksi ciri-ciri visual tertentu, seperti logo dan stempel yang menjadi indikator tidak kredibel.

Salah satu pendekatan yang dapat digunakan untuk mendukung proses verifikasi sertifikat adalah teknologi *computer vision*. Teknologi ini memungkinkan sistem untuk mengenali pola, objek, atau elemen visual tertentu secara otomatis dari citra digital, sehingga tidak lagi bergantung pada pemeriksaan manual [10]. Metode *Scale Invariant Feature Transform* (SIFT) merupakan salah satu teknik ekstraksi fitur dalam *computer vision* yang mampu menemukan *keypoint* atau ciri unik dari sebuah citra, yang bersifat invarian terhadap skala, rotasi, dan pencahayaan [11]. Penelitian [12] membahas implementasi algoritma SIFT dalam sistem pendeteksi logo untuk keperluan verifikasi dokumen. Dalam penelitian tersebut, dilakukan proses segmentasi berbasis ambang dan deteksi tepi, serta ekstraksi wilayah calon logo menggunakan metode *Maximally Stable Extremal Regions* (MSER). Setelah area kandidat logo diperoleh, fitur-fitur lokal diekstraksi menggunakan algoritma SIFT untuk menghasilkan deskriptor yang mewakili ciri visual dari setiap logo. Kemudian dibandingkan dengan basis data logo institusi melalui proses *feature matching* dan penyaringan pasangan untuk menentukan kecocokan logo yang valid. Hasil penelitian menunjukkan bahwa kombinasi antara segmentasi otomatis dan ekstraksi fitur SIFT mampu mendeteksi dan mengenali logo dengan tingkat akurasi yang tinggi pada citra dokumen uji [12]. Implementasi SIFT terbukti efektif dalam membedakan logo antar institusi meskipun terdapat variasi ukuran, rotasi, maupun pencahayaan pada citra masukan. Temuan tersebut menunjukkan bahwa SIFT memiliki potensi yang baik untuk digunakan dalam sistem verifikasi dokumen berbasis citra karena mampu mendeteksi dan mencocokkan fitur visual secara konsisten.

Namun demikian, metode SIFT standar memiliki beberapa kelemahan, seperti komputasi yang lambat karena memiliki ukuran deskriptor yang besar (128 dimensi), serta sensitivitas terhadap perubahan pencahayaan ekstrem [13]. Kondisi ini dapat mengakibatkan *false match* atau kesalahan pencocokan fitur pada citra

dengan karakter visual yang mirip namun tidak relevan. Untuk mengatasi kekurangan tersebut, berbagai penelitian kemudian mengembangkan *Improved SIFT* dengan tujuan meningkatkan efisiensi dan stabilitas deteksi fitur. Penelitian [13] berfokus pada peningkatan efisiensi pencocokan citra melalui pengembangan stabilitas fitur dan penyederhanaan deskriptor dari 128 menjadi 96 dimensi, yang terbukti mampu mempercepat proses pencocokan dan meningkatkan akurasi hasil pencocokan citra. Uji eksperimental menunjukkan bahwa metode ini mempercepat waktu pencocokan hingga 30% dan membuktikan bahwa deskriptor berukuran lebih kecil tetap dapat menghasilkan pencocokan fitur yang presisi dan efisien. Sementara itu, Penelitian [14] menerapkan konsep *Improved SIFT* pada sistem identifikasi sidik jari untuk meningkatkan ketahanan terhadap *noise* dan kontras rendah. Dalam penelitian tersebut, algoritma SIFT diadaptasi dengan proses *adaptive preprocessing* (meliputi *denoising*, *binarization*, *thinning*, dan *dilation*) serta penggunaan *dual-threshold filtering* untuk meminimalkan kecocokan palsu pada citra minutiae dapat meningkatkan ketahanan terhadap *noise* dan kualitas citra yang rendah. Temuan - temuan tersebut menunjukkan bahwa *Improved SIFT* memiliki potensi untuk menghasilkan proses pencocokan fitur yang lebih efisien, stabil, dan andal dibandingkan SIFT standar.

Akan tetapi, hasil pencocokan fitur dari SIFT masih berpotensi menghasilkan kesalahan identifikasi berupa *false match*, khususnya pada citra dengan pola visual mirip namun tidak relevan [15]. Untuk mengatasi hal ini, algoritma *Random Sample Consensus* (RANSAC) digunakan untuk menyaring pasangan fitur yang tidak konsisten dan hanya mempertahankan yang sesuai dengan homografi [16]. Penelitian [17] mengenai *Logo Matching for Document Image Retrieval Using SIFT* membahas penerapan algoritma SIFT dalam proses pencocokan logo pada citra dokumen dengan menambahkan pendekatan berbasis konteks spasial untuk meningkatkan keandalan deteksi. Dalam penelitian tersebut, fitur-fitur SIFT diekstraksi dari citra dokumen dan dibandingkan dengan logo acuan. Selanjutnya, RANSAC digunakan untuk memvalidasi kesesuaian geometris antar pasangan titik fitur. Hasil eksperimen menunjukkan bahwa integrasi antara SIFT dan RANSAC, mampu meningkatkan akurasi pencocokan logo secara

signifikan dibandingkan dengan metode SIFT standar. Pendekatan ini tidak hanya mengurangi jumlah *false match*, tetapi juga mempercepat proses pengenalan logo dalam citra dokumen, menjadikannya solusi efektif untuk aplikasi *document image retrieval* berbasis kesamaan visual [17].

Meskipun berbagai penelitian telah menerapkan SIFT dan *Improved SIFT* pada pengenalan logo maupun identifikasi biometrik, penerapan kombinasi *Improved SIFT* dan RANSAC untuk verifikasi sertifikat prestasi pada proses seleksi SNBP masih relatif terbatas. Berdasarkan berbagai penelitian terdahulu, algoritma SIFT terbukti dapat mengekstraksi ciri visual yang invarian terhadap perubahan skala, rotasi, dan pencahayaan, serta banyak digunakan pada bidang pengenalan objek visual yang memiliki pola unik. Namun, penerapan SIFT standar masih menghadapi beberapa keterbatasan, seperti waktu komputasi yang tinggi dan kemungkinan terjadinya *false match* pada citra dengan pola visual yang mirip namun tidak relevan. Dalam konteks verifikasi sertifikat prestasi SNBP yang jumlahnya mencapai jutaan dan memiliki variasi desain, kualitas, serta resolusi yang beragam, dibutuhkan pendekatan yang efisien tanpa mengorbankan akurasi. Untuk menjawab tantangan ini, penelitian ini menerapkan adaptasi *Improved Scale Invariant Feature Transform* yang mengoptimalkan stabilitas dan efisiensi deteksi fitur melalui proses *stability filtering* pada *keypoints* serta penyederhanaan deskriptor dari 128 menjadi 96 dimensi. Metode ini dikombinasikan dengan *Random Sample Consensus* (RANSAC) untuk memvalidasi hubungan geometris antar fitur dan menghapus pasangan fitur yang tidak konsisten.

Selain itu, Penelitian [18] mengembangkan sistem pengenalan QR Code secara *real time* menggunakan *OpenCV* dan pustaka *pyzbar* untuk mendeteksi serta mendekode informasi yang tersimpan pada QR Code. Sistem diawali dengan proses *image preprocessing*, dilanjutkan dengan deteksi lokasi QR Code, proses *decoding*, kemudian menampilkan informasi hasil ekstraksi berupa URL atau data teks. Hasil penelitian menunjukkan bahwa sistem mampu mendeteksi dan mendekode QR Code dengan tingkat akurasi sebesar 95,5%, sehingga dinilai efektif untuk proses autentikasi dan pengambilan informasi secara otomatis. Berdasarkan keberhasilan tersebut, penelitian ini mengadopsi deteksi QR Code sebagai fitur pendukung

identifikasi. URL yang berhasil diekstraksi dari *QR Code* digunakan untuk mengakses halaman verifikasi, kemudian informasi pada halaman tersebut dibandingkan dengan isi sertifikat menggunakan OCR sehingga dapat memberikan bukti tambahan terhadap hasil identifikasi berdasarkan pencocokan logo dan stempel. Selain *QR Code*, Penelitian [19] mengimplementasikan *Optical Character Recognition* (OCR) menggunakan *Tesseract* OCR untuk mengekstraksi karakter dari citra teks. Penelitian tersebut menerapkan tahapan *pre-processing*, pengenalan karakter, dan *post-processing* sehingga teks pada citra dapat diubah menjadi data digital secara otomatis. Hasil penelitian menunjukkan bahwa *Tesseract* OCR mampu mempercepat proses ekstraksi teks dibandingkan pengetikan manual dan memberikan hasil yang baik apabila didukung oleh kualitas citra yang memadai. Dari penelitian tersebut, OCR pada penelitian ini dimanfaatkan untuk mengekstraksi informasi teks pada sertifikat, baik untuk membandingkan isi sertifikat dengan halaman hasil *QR Code* maupun untuk mengidentifikasi tingkatan prestasi, seperti kabupaten, kota, provinsi, nasional, dan internasional, sehingga proses identifikasi dapat dilakukan secara lebih otomatis.

Penelitian ini menggunakan dataset 200 sertifikat prestasi calon mahasiswa jalur SNBP yang diperoleh dari panitia penyelenggara, mencakup variasi format file (PDF, JPG, dan PNG), perbedaan desain, kualitas pemindaian, serta keberadaan maupun ketiadaan *QR Code*. Variasi data tersebut dipilih untuk merepresentasikan kondisi nyata dalam proses seleksi SNBP yang beragam dan kompleks, sehingga sistem yang dibangun dapat diuji terhadap berbagai karakteristik dokumen. Berdasarkan permasalahan tersebut, penelitian ini berkontribusi dalam penerapan metode *Improved SIFT* dengan deskriptor 96 dimensi untuk verifikasi visual sertifikat, integrasi algoritma RANSAC untuk meningkatkan keandalan pencocokan fitur melalui validasi geometris, serta penambahan fitur pendukung berupa deteksi *QR Code* dan analisis OCR untuk memperkaya informasi hasil verifikasi secara otomatis. Dengan pendekatan ini, penelitian diharapkan dapat menghasilkan sistem verifikasi sertifikat yang lebih cepat, akurat, dan andal, serta mampu meningkatkan transparansi dan objektivitas dalam proses seleksi jalur prestasi di perguruan tinggi.

1.2. Rumusan Masalah

Rumusan masalah yang menjadi fokus utama pada penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan tahapan *preprocessing* untuk meningkatkan kualitas citra sertifikat?
2. Bagaimana cara mengekstraksi dan mencocokkan fitur visual pada citra sertifikat untuk mendapatkan pola yang relevan?
3. Bagaimana merancang antarmuka pengguna (GUI) yang memudahkan pengguna dalam mengunggah sertifikat dan memperoleh hasil identifikasi otomatis?

1.3. Batasan Masalah

Batasan yang ditetapkan pada masalah yang diamati pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa kumpulan sertifikat prestasi jalur SNBP dalam format PDF, JPG, dan PNG, yang diperoleh dari dokumen peserta seleksi, diproses sebagai citra raster,
2. Proses identifikasi pada penelitian ini hanya difokuskan pada pencocokan ciri visual berupa logo dan stempel sebagai parameter identifikasi, tanpa mempertimbangkan informasi yang terkandung dalam isi teks sertifikat.
3. Informasi yang diperoleh dari *QR Code* dan OCR tingkat capaian prestasi hanya digunakan sebagai informasi pendukung.
4. Penelitian ini tidak mencakup pendekatan berbasis *deep learning*, dan seluruh proses analisis dilakukan menggunakan metode berbasis pengolahan citra digital tanpa pelatihan model.
5. Aplikasi yang dikembangkan berupa aplikasi desktop yang berjalan secara lokal, dan belum diimplementasikan pada platform berbasis web maupun terintegrasi dengan sistem SNBP.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menjawab rumusan masalah serta memberikan arahan yang lebih jelas bagi penelitian ini. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Menggunakan teknik peningkatan kontras, konversi citra ke *grayscale*, dan penyesuaian ukuran citra untuk meningkatkan kualitas visual sertifikat agar hasil menjadi lebih optimal.
2. Mengimplementasikan metode *Improved SIFT* dan *RANSAC* untuk mengekstraksi, mencocokkan, dan memvalidasi fitur visual pada citra sertifikat prestasi jalur SNBP.
3. Membuat aplikasi desktop berbasis *PySide* yang dapat mengunggah sertifikat, menjalankan proses identifikasi secara otomatis, dan menampilkan hasil identifikasi secara langsung kepada pengguna

1.5. Manfaat Penelitian

Manfaat penelitian yang dapat diperoleh dari hasil penelitian ini adalah sebagai berikut:

a. Manfaat Teoritis

Penelitian ini dapat memperkaya literatur mengenai penerapan metode *computer vision*, khususnya integrasi *Improved SIFT*, dan *RANSAC*, dalam identifikasi dokumen pendidikan. Selain itu, penelitian ini memberikan kontribusi pada pengembangan pengetahuan tentang bagaimana teknik ekstraksi fitur visual dapat digunakan untuk menilai kredibilitas sertifikat prestasi mahasiswa

b. Manfaat Praktis

1. Bagi Panitia SNBP: penelitian ini menyediakan solusi identifikasi sertifikat prestasi berbasis visual yang lebih cepat, akurat, dan konsisten, sehingga mampu mengurangi ketergantungan pada proses manual yang memakan waktu.

2. Bagi Penulis: penelitian ini menjadi sarana untuk meningkatkan keterampilan dalam bidang *computer vision*, pengolahan citra digital, serta pengembangan aplikasi dengan antarmuka pengguna yang praktis.
3. Bagi Peneliti Selanjutnya: penelitian ini dapat menjadi acuan dalam pengembangan sistem identifikasi dokumen berbasis fitur visual untuk konteks lain, misalnya dokumen identitas atau ijazah, sehingga memperluas potensi penerapan teknologi ini di berbagai bidang