

AUTOMATIC IMAGE ANNOTATION MENGGUNAKAN METODE SPEEDED UP ROBUST FEATURE DAN SUPPORT VECTOR MACHINE

AUTOMATIC IMAGE ANNOTATION USING SPEEDED UP ROBUST FEATURES AND SUPPORT VECTOR MACHINE METHOD

Nurul Halimatul Azizah¹, Febryanti Sthevanie ST., MT.², Anditya Arifianto ST., MT³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

¹nurhalizahh@telkomuniversity.ac.id, ²ssthevanie@telkomuniversity.ac.id, ³anditya@telkomuniversity.ac.id

Abstrak

Tersedianya citra dalam jumlah yang cukup banyak menyebabkan dibutuhkan sistem untuk mengorganisasikan citra sesuai dengan kategori tertentu untuk memudahkan user untuk mencari citra yang dibutuhkan. Pencarian citra menggunakan teks saat ini masih kurang efektif karena pemberian teks terhadap citra oleh masing-masing orang masih terlalu subyektif dan memakan waktu yang cukup lama. Pencarian dengan menggunakan citra sebagai query menggunakan fitur yang terdapat pada citra juga masih kurang efektif karena membutuhkan citra sebagai query dan hasilnya masing kurang sesuai. Sistem Automatic Image Annotation merupakan sistem yang dapat mengatasi kekurangan yang terdapat pada pencarian citra menggunakan teks dan pencarian citra menggunakan citra query, dengan cara memberikan anotasi secara otomatis terhadap citra menggunakan fitur yang terdapat pada citra. Dalam penelitian tugas akhir ini, akan dibangun sebuah sistem Automatic Image Annotation menggunakan metode ekstraksi ciri Speeded Up Robust Features (SURF) dan metode klasifikasi Support Vector Machine (SVM) menggunakan Bag-of-Visual-Words. Dari hasil pengujian, akurasi tertinggi didapat menggunakan nilai threshold 0, jumlah kluster 500, dan menggunakan jenis SVM *one-against-all* dengan hasil F1 Score rata-rata sebesar 0,9001.

Kata kunci : Automatic Image Annotation, SURF, SVM, *visual words*.

Abstract

Due to the availability of images in a large scale nowadays, a system to organize images according to certain categories is needed to help user to find a required image. Image search using text is no longer effective as captioning the image by each person is too subjective and takes a long time. Image search with image as a query using image features is also no longer effective as it need the image as a query and the results are still not match. Automatic Image Annotation is a system that can overcome the lack in image search using text and image retrieval using query images, by giving an automatic annotation to the image using image features. In this research, Automatic Image Annotation system will be built using Speeded Up Robust Features (SURF) as feature extraction method and Support Vector Machine using Bag-of-Visual-Words as classification method. From the experiments, the highest F1 Score is 0,9001 obtained with 0 threshold, 500 clusters, and one-against-all SVM.

Keywords : Automatic Image Annotation, SURF, SVM, *visual words*.

1. Pendahuluan

Seiring dengan berkembangnya teknologi digital saat ini, informasi dalam bentuk citra sudah tersedia dengan cukup banyak. Dalam pencarian informasi berupa citra atau yang biasa disebut dengan *Image Retrieval* (IR), cara tradisional yang biasa digunakan adalah pencarian menggunakan teks yang sudah diberikan kepada masing-masing citra yang disebut dengan anotasi secara manual sebelumnya, atau biasa disebut dengan *Text Based Image Retrieval* [9], namun, pemberian anotasi kepada suatu citra dapat berbeda-beda antara satu orang dengan yang lainnya, sehingga dinilai terlalu subyektif [1]. Pemberian anotasi kepada masing-masing citra secara manual juga membutuhkan waktu yang cukup lama sehingga dibutuhkan sistem yang lebih efisien [4], oleh karena itu, saat ini berkembang sistem yang dapat melakukan pencarian citra melalui fitur yang terdapat pada citra tersebut, yaitu *Content Based Image Retrieval* (CBIR). CBIR merupakan sebuah sistem untuk pencarian citra dimana user memasukkan sebuah citra sebagai *query*, kemudian sistem akan memberikan hasil beberapa citra yang relevan terhadap citra *query* tersebut berdasarkan ciri yang terdapat pada citra, contohnya seperti warna atau tekstur [4]. Masih terdapat kekurangan dalam sistem ini, yaitu citra yang dihasilkan terkadang berbeda dari hasil interpretasi manusia terhadap citra tersebut, karena banyak citra yang memiliki ciri yang sama namun berasal dari kategori yang berbeda. Perbedaan tersebut disebut dengan *semantic gap* [1]. Hal tersebut dapat diatasi dengan *Automatic*

Image Annotation (AIA). AIA merupakan sebuah sistem untuk memberikan anotasi terhadap citra secara otomatis berdasarkan ciri yang ada pada citra tersebut dan dengan mempelajari citra yang sudah diberikan anotasi sebelumnya, sehingga sistem dapat menentukan kategori dari citra dan melakukan penelusuran berdasarkan kategori tersebut.

Beberapa penelitian mengenai AIA sebelumnya menggunakan metode ekstraksi fitur global yaitu warna, tekstur, dan bentuk [1], kelemahan dalam penggunaan fitur global ini adalah pada warna terdapat banyak citra yang berbeda secara visual namun memiliki penggunaan warna yang sama. Pada tekstur, terdapat beberapa metode yang tidak dapat mengenali citra yang mengalami perubahan skala dan rotasi. Sedangkan pada bentuk, penggunaan ekstraksi ciri bentuk saja masih kurang handal sehingga biasanya dikombinasikan dengan metode ekstraksi ciri lain agar lebih efektif.

Penelitian lainnya menggunakan fitur lokal untuk ekstraksi ciri. Penggunaan fitur lokal terbukti memberikan akurasi yang lebih tinggi pada *Image Retrieval* [14][15]. Beberapa contoh metode fitur lokal adalah *Scale Invariant Feature Transform* (SIFT) [16] dan *Speeded Up Robust Features* (SURF) [8] yang menghasilkan ciri berupa *keypoint* dari citra. Penggunaan SURF terbukti dapat memberikan hasil yang lebih baik dibandingkan dengan SIFT [8].

Pada penelitian kali ini, metode ekstraksi ciri yang digunakan adalah *Speeded Up Robust Features* (SURF) dan metode klasifikasi yang digunakan adalah *Support Vector Machine* (SVM). Metode SURF menggunakan fitur lokal untuk ekstraksi ciri. SURF telah terbukti dapat mengenali citra yang mengalami perubahan skala, rotasi, *blurring*, pencahayaan, dan perubahan sudut pandang terhadap objek yang ada pada citra [8]. Sedangkan metode SVM terbukti dapat menangani masalah klasifikasi pada data yang memiliki dimensi fitur yang tinggi dengan cukup baik [10].

2. Landasan Teori

2.1. Speeded Up Robust Features (SURF)

Speeded Up Robust Features (SURF) merupakan sebuah *detector* dan *descriptor* ciri dari citra dengan mencari *keypoint* pada citra. Pencarian *keypoint* dilakukan dengan melihat citra secara keseluruhan tanpa perlu melakukan segmentasi, sehingga lebih cepat dan lebih akurat. SURF telah terbukti dapat mengenali citra yang mengalami perubahan skala, rotasi, *blurring*, pencahayaan, dan perubahan sudut pandang terhadap objek yang ada pada citra [8].

SURF mencari *keypoint* berdasarkan matriks Hessian. Untuk titik $X = (x, y)$ pada citra I , matriks Hessian $H(X, \sigma)$ pada posisi X dengan skala σ didefinisikan dengan :

$$H(X, \sigma) = \begin{pmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{pmatrix} \quad (1)$$

Dimana $L_{xx}(X, \sigma)$ merupakan hasil konvolusi dari turunan kedua fungsi gaussian $\frac{\partial^2}{\partial x^2} g(\sigma)$ dengan citra I pada titik X , demikian pula dengan $L_{yy}(X, \sigma)$ yang merupakan hasil konvolusi dari turunan kedua fungsi gaussian $\frac{\partial^2}{\partial y^2} g(\sigma)$ dan $L_{xy}(X, \sigma)$ yang merupakan hasil konvolusi dari turunan kedua fungsi gaussian $\frac{\partial^2}{\partial x \partial y} g(\sigma)$.

Untuk mempercepat komputasi, turunan kedua fungsi gaussian untuk matriks Hessian dihitung menggunakan *box filter* yang telah didiskritisasi. Sehingga hasil konvolusi dari turunan kedua fungsi gaussian menggunakan *box filter* yang telah didiskritisasi dinotasikan sebagai $D_{xx}(X, \sigma)$, $D_{yy}(X, \sigma)$, dan $D_{xy}(X, \sigma)$. Pencarian interest point kemudian dilakukan menggunakan nilai determinan yang dirumuskan dengan :

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2)$$

Interest point kemudian dideteksi pada ruang skala dan oktaf yang berbeda-beda dengan cara memperbesar skala dari filter. Jarak dari *interest point* disimpan dalam $3 \times 3 \times 3$ ketetanggaan untuk kemudian dilokalisasi menggunakan penekanan non-maksimum terhadap nilai determinan.

Keypoint descriptor dari citra kemudian ditentukan dengan menentukan orientasi dari *interest point* dengan menghitung respon Haar Wavelet dari arah x dan y . Setelah itu, daerah persegi di sekitar *interest point* searah dengan orientasinya diekstrak dengan membagi daerah tersebut menjadi 4×4 sub bagian kemudian dilakukan penghitungan *gradient histogram* yang didefinisikan dengan :

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right) \quad (3)$$

Dimana d_x adalah respon Haar Wavelet pada arah horisontal dan d_y adalah arah vertikal. Vektor fitur dari SURF didapatkan dengan menggabungkan fitur dari sub bagian *keypoint* sehingga menghasilkan vektor dengan 64 elemen.

2.2. Support Vector Machine (SVM)

Support Vector Machine merupakan salah satu metode klasifikasi *supervised learning* yang bekerja dengan cara mencari *hyper plane* atau bidang pembatas yang memisahkan data menjadi dua kelas dengan jarak maksimum diantara masing-masing kelas [4]. Untuk data dengan jumlah n sampel, $\{x_i, y_i\}$, *hyperplane* pemisah dapat didefinisikan dengan memenuhi :

$$y_i(w \cdot x_i + b) > 0 \quad \forall i = 1, 2, \dots, N \quad (4)$$

Dimana $x_i \in R^d$ merupakan vektor dari fitur dengan d dimensi dan $y_i \in \{+1, -1\}$ merupakan label yang mengindikasikan kelas. Data dapat dikatakan terpisah secara linear jika terdapat *hyperplane* yang dapat memisahkannya. SVM kemudian memilih jarak ke kelas terdekat dari *hyperplane* dengan $\frac{1}{\|w\|}$ sehingga :

$$y_i(w \cdot x_i + b) \geq 1 \quad (5)$$

Hyperplane yang optimal didapatkan dengan meminimasi $\|w^2\|$. Margin didefinisikan dengan $\frac{2}{\|w\|}$ dan data terdekat dengan *hyperplane* disebut dengan *support vectors* [10]. Pencarian margin maksimum dapat dilakukan dengan meminimasi persamaan (6) yang memenuhi (7).

$$J(w) = \frac{1}{2} \|w\|^2 \quad (6)$$

$$z_i(w^t x_i + w_o) \geq 1 \quad \forall i \quad (7)$$

$J(w)$ merupakan sebuah fungsi kuadrat, sehingga memiliki satu nilai global minimum. Permasalahan tersebut dapat diselesaikan menggunakan *Lagrange Multiplier* yang dirumuskan dengan memaksimalkan persamaan (8) yang memenuhi (9) dan (10).

$$L_D(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j z_i z_j x_i^t x_j \quad (8)$$

$$a_i \geq 0 \quad \forall i \quad (9)$$

$$\sum_{i=1}^n a_i z_i = 0 \quad (10)$$

$L_D(a)$ dapat dioptimasi menggunakan Quadratic Programming dan nilai $a = \{a_1, \dots, a_n\}$ merupakan variabel baru sejumlah data latih.

- Apabila nilai $a_i = 1$ maka data i bukan merupakan support vector.
- Apabila nilai $a_i \neq 0$ dan $z_i(w^t x_i + w_o - 1) = 0$ maka data i merupakan support vector.

Nilai w dapat dicari menggunakan persamaan $w = \sum_{i=1}^n a_i z_i x_i$ dan nilai w_o dapat dicari menggunakan salah satu nilai yang memenuhi $a_i > 0$ dan $a_i [z_i(w^t x_i + w_o) - 1] = 0$ sehingga $w_o = \frac{1}{z_i} - w^t x_i$.

Untuk kasus data yang tidak bisa dipisahkan secara linear, pencarian *hyperplane* dapat dilakukan dengan *soft margin SVM*, yaitu dengan menggunakan variabel *slack* ξ untuk masing-masing data. Sehingga persamaan (7) dapat dimodifikasi menjadi :

$$z_i(w^t x_i + w_o) \geq 1 - \xi_i \quad \forall i \quad (11)$$

- Apabila $\xi_i > 1$ maka data i berada dalam sisi yang salah dari *hyperplane*.
- Apabila $0 < \xi_i < 1$ maka data i berada dalam sisi yang benar dari *hyperplane* namun berada di dalam margin.
- Apabila $\xi_i < 0$ maka data i berada dalam posisi yang ideal.

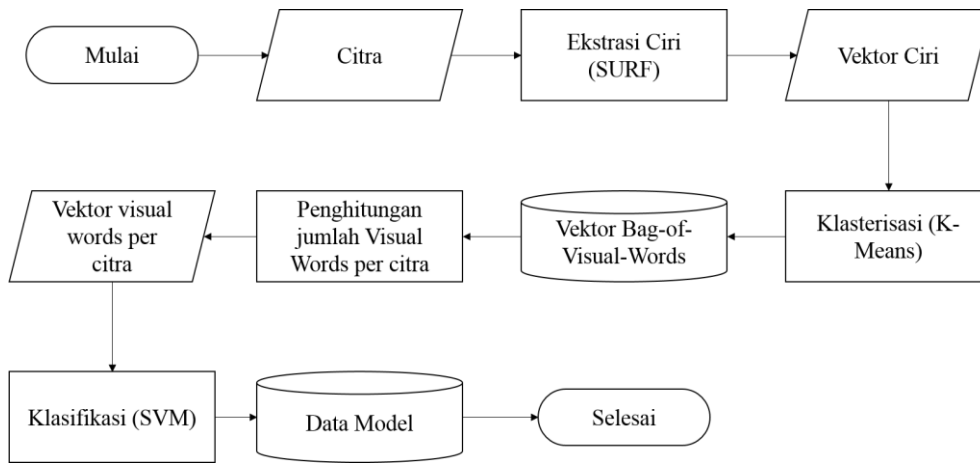
Kemudian pencarian nilai a dapat menggunakan persamaan (8) yang memenuhi (12) dan (10).

$$0 \leq a_i \leq \beta \quad \forall i \tag{12}$$

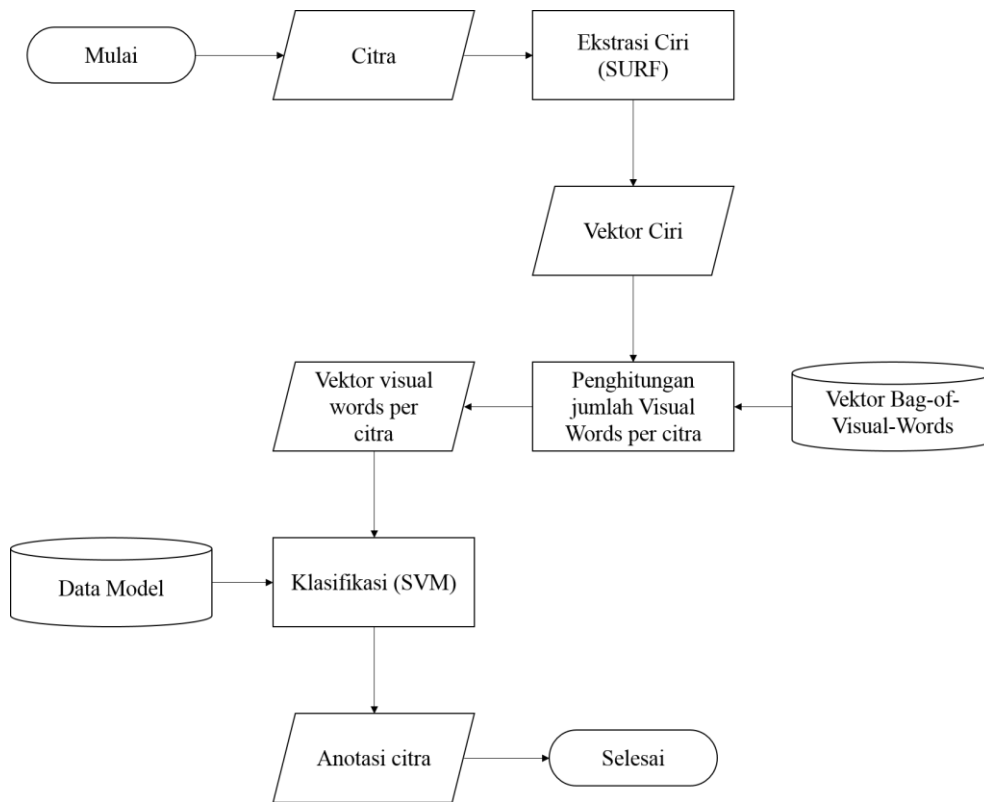
β merupakan nilai konstan yang apabila bernilai kecil maka ada banyak sampel yang dibiarkan tidak berada dalam posisi ideal, sedangkan apabila bernilai besar maka hanya ada sedikit sampel yang tidak berada dalam posisi ideal.

3. Perancangan Sistem

Tahapan pertama yang dilakukan dalam sistem ini adalah melakukan ekstraksi ciri menggunakan SURF untuk mendapat fitur lokal *keypoint* dari citra. Kemudian dilakukan proses klasterisasi menggunakan K-Means terhadap *keypoint* tersebut untuk membentuk *Bag-of-Visual-Words*. *Bag-of-Visual-Words* merupakan model yang diadaptasi dari model *Bag-of-Words* yang digunakan dalam pemrosesan teks [12]. Proses klasterisasi dilakukan untuk mengelompokkan fitur dari citra yang memiliki ciri-ciri yang sama untuk membentuk *visual words*, sehingga masing-masing citra dapat direpresentasikan dengan *visual words* yang terdapat di dalamnya [11]. Model klasifikasi SVM kemudian dibangun menggunakan *visual words* dari citra. Tahapan yang akan dilakukan dalam perancangan sistem ini terbagi menjadi dua tahapan, yaitu proses pelatihan (Gambar 1) dan proses anotasi (Gambar 2).



Gambar 1 Alur proses pelatihan sistem



Gambar 2 Alur proses pengujian sistem

4. Hasil dan Analisis Pengujian

Terdapat beberapa pengujian yang dilakukan di penelitian ini. Pengujian dilakukan untuk memperoleh parameter yang optimal sehingga dapat menghasilkan akurasi terbaik yang diukur menggunakan *F1-Measure*. Data citra yang digunakan adalah 10 kelas citra yang terdiri dari 5 kelas citra satu buah objek yang diambil dari dataset Caltech101 [17] dan lima kelas citra pemandangan yang diambil dari dataset Corel database [18], kedua dataset tersebut memiliki ukuran citra yang berbeda-beda. Kategori dari 10 kelas citra tersebut diantaranya : *airplanes, faces, ketch, leopards, motorbikes, fireworks, indoor, night, rural, dan sunset*.

Sejumlah 100 citra dari dataset tersebut kemudian dibagi menjadi 60 citra untuk data latih dan 40 citra untuk data uji. Karena terdapat unsur random dalam proses klusterisasi menggunakan K-Means sehingga parameter yang sama dapat menghasilkan hasil yang berbeda-beda, maka pengujian terhadap setiap parameter dilakukan sebanyak 20 kali, dan akurasi parameter tersebut adalah rata-rata dari 20 kali pengujian. Hasil centroid untuk visual words final kemudian ditentukan berdasarkan parameter yang memiliki rata-rata akurasi tertinggi dan nilai akurasi tertinggi dari 20 kali pengujian yang dilakukan. Pengujian dilakukan menggunakan *soft margin SVM*, karena selain dapat mengatasi data yang tidak dapat terpisah secara linear, *soft margin SVM* juga dapat mengatasi data yang terdapat *noise* dan *outliers* sehingga menghasilkan model yang memiliki generalisasi yang baik [20]. Terdapat 2 buah skenario yang dilakukan pada sistem ini. Penjelasan mengenai skenario yang diuji dapat dilihat pada Tabel 1.

Tabel 1 Skenario pengujian

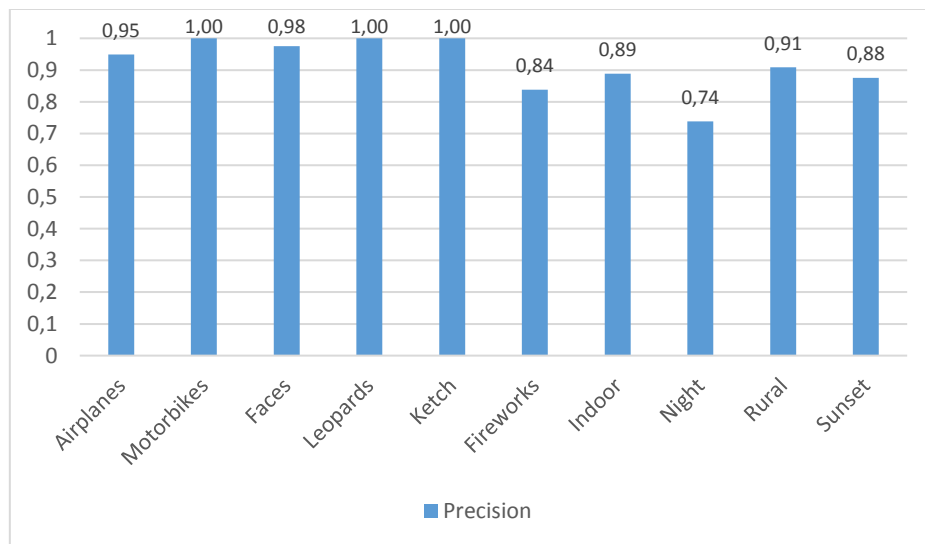
Skenario	Tujuan	Parameter Uji	Parameter Konstan
1	Mengetahui pengaruh nilai threshold keypoint dan jumlah klaster terhadap akurasi sistem.	Threshold = [0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]; Klaster = [200, 300, 400, 500]	SVM = <i>one-against-all</i> ;
2	Mengetahui pengaruh jenis SVM yang digunakan terhadap akurasi sistem.	SVM = [<i>one-against-all, pairwise SVM</i>]	Hasil Skenario-1

Dari kedua skenario pengujian yang dilakukan, hasil terbaik yang memberikan akurasi tertinggi didapat dengan menggunakan nilai threshold = 0, jumlah klaster = 500, dan SVM one-against-all. Hasil centroid visual words final kemudian diambil dari hasil pengujian menggunakan parameter-parameter tersebut. Hasil pengujian terbaik didapat pada pengujian ke 7 dari 20 kali pengujian dengan nilai akurasi 0,9149. Hasil confusion matrix dari hasil tersebut dapat dilihat pada tabel 2.

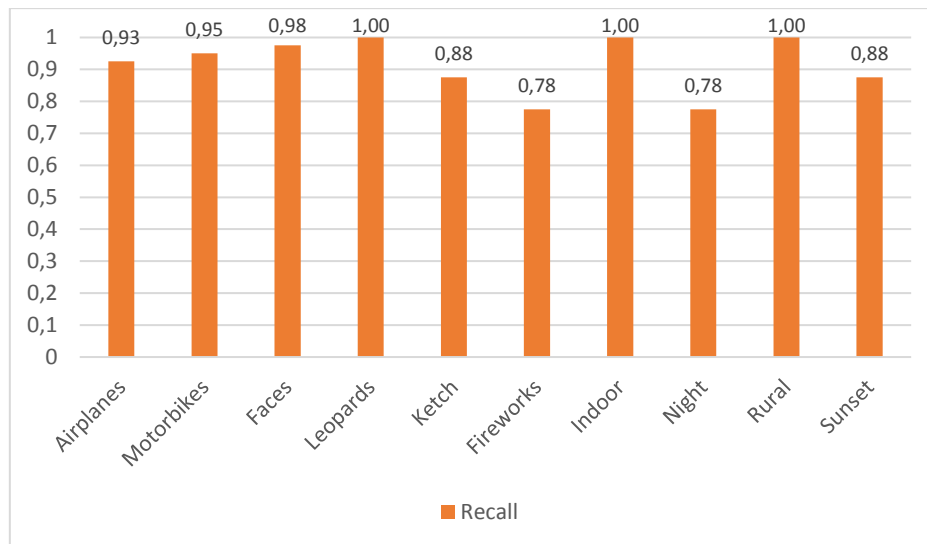
Tabel 2 Confusion matrix hasil pengujian terbaik

		Kelas hasil prediksi									
		Airplanes	Motorbikes	Faces	Leopards	Ketch	Fireworks	Indoor	Night	Rural	Sunset
Kelas seharusnya	Airplanes	37	0	1	0	0	0	2	0	0	0
	Motorbikes	1	38	0	0	0	0	1	0	0	0
	Faces	0	0	39	0	0	0	1	0	0	0
	Leopards	0	0	0	40	0	0	0	0	0	0
	Ketch	1	0	0	0	35	0	0	0	2	2
	Fireworks	0	0	0	0	0	31	0	7	2	0
	Indoor	0	0	0	0	0	0	40	0	0	0
	Night	0	0	0	0	0	5	1	31	0	3
	Rural	0	0	0	0	0	0	0	0	40	0
	Sunset	0	0	0	0	0	1	0	4	0	35

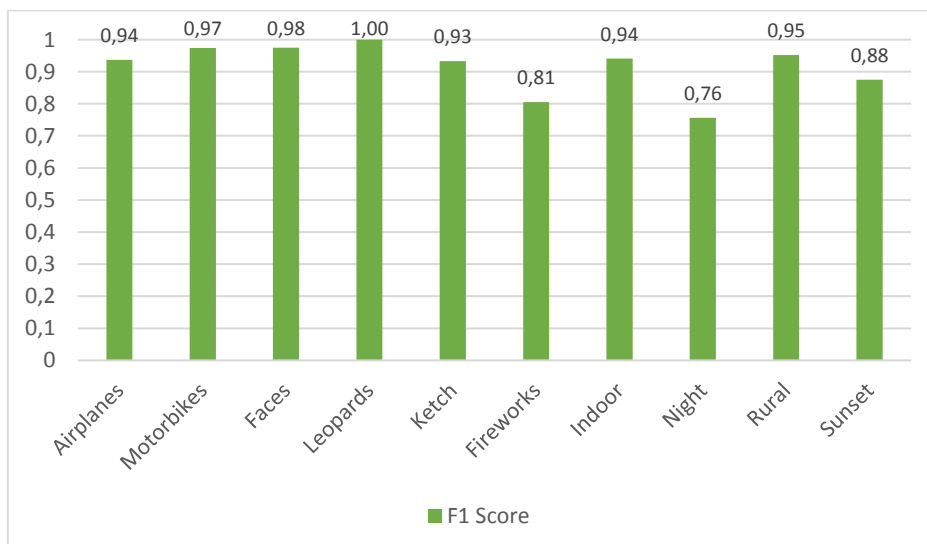
Sedangkan hasil precision, recall, dan F1 Score dari masing-masing kelas dapat dilihat pada gambar 3, gambar 4, dan gambar 5.



Gambar 3 Precision masing-masing kelas hasil pengujian terbaik



Gambar 4 Recall masing-masing kelas hasil pengujian terbaik



Gambar 5 F1 Score masing-masing kelas hasil pengujian terbaik

5. Kesimpulan

Berdasarkan proses implementasi dan analisis dari hasil pengujian yang dilakukan dalam tugas akhir ini, dapat ditarik kesimpulan bahwa metode ekstraksi ciri Speeded Up Robust Features dan metode klasifikasi Support Vector Machine dapat memberikan anotasi terhadap citra dengan baik yang menghasilkan nilai F1 Score rata-rata sebesar 0,9001 menggunakan parameter nilai threshold 0, jumlah kluster 500, dan jenis SVM one-against-all. Thresholding terhadap keypoint pada Speeded Up Robust Features menurunkan akurasi sistem dan semakin tinggi jumlah kluster, akurasi yang dihasilkan semakin tinggi.

Daftar pustaka :

- [1] ZHANG, Dengsheng; ISLAM, Md Monirul; LU, Guojun. A review on automatic image annotation techniques. *Pattern Recognition*, 2012, 45.1: 346-362.
- [2] SANGEETHA, M.; ANANDAKUMAR, K.; BHARATHI, A. Automatic Image Annotation and Retrieval: A Survey. 2016.
- [3] ANEES, V. Muhammed; KUMAR, G. Santhosh; SREERAJ, M. Automatic image annotation using SURF descriptors. In: *2012 Annual IEEE India Conference (INDICON)*. IEEE, 2012. p. 920-924.
- [4] SHUKLA, Tuhin; MISHRA, Nishchol; SHARMA, Sanjeev. Automatic image annotation using SURF features. *International Journal of Computer Applications*, 2013, 68.4.

- [5] PAUL, Edwin; AS, Ajeena Beegom. Mining images for image annotation using SURF detection technique. In: *2015 International Conference on Control Communication & Computing India (ICCC)*. IEEE, 2015. p. 724-728.
- [6] DA SILVA TORRES, Ricardo; FALCAO, Alexandre X. Content-Based Image Retrieval: Theory and Applications. *RITA*, 2006, 13.2: 161-185.
- [7] DATTA, Ritendra, et al. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 2008, 40.2: 5.
- [8] BAY, Herbert, et al. Speeded-up robust features (SURF). *Computer vision and image understanding*, 2008, 110.3: 346-359.
- [9] RUI, Yong; HUANG, Thomas S.; CHANG, Shih-Fu. Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, 1999, 10.1: 39-62.
- [10] CHAPELLE, Olivier; HAFFNER, Patrick; VAPNIK, Vladimir N. Support vector machines for histogram-based image classification. *IEEE transactions on Neural Networks*, 1999, 10.5: 1055-1064.
- [11] HWANG, Sae. Bag-Of-Visual-Words Approach based on SURF Features to Polyp Detection in Wireless Capsule Endoscopy Videos. *Proceeding ISVC*, 11: 320-327.
- [12] SHEKHAR, Ravi; JAWAHAR, C. V. Word image retrieval using bag of visual words. In: *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*. IEEE, 2012. p. 297-301.
- [13] Abe, Shigeo. "Analysis of multiclass support vector machines." *Thyroid* 21.3 (2003): 3772.
- [14] ALY, Mohamed, et al. Automatic discovery of image families: Global vs. local features. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009. p. 777-780.
- [15] BANNOUR, Hichem; HLAOUA, Lobna; EL AYEB, Béchir. Survey of the Adequate Descriptor for Content-Based Image Retrieval on the Web: Global versus Local Features. In: *CORIA*. 2009. p. 445-456.
- [16] LOWE, David G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 2004, 60.2: 91-110.
- [17] L. Fei-Fei, R. Fergus and P. Perona. *Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories*. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004
- [18] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 9, pp. 947-963, September 2001.
- [19] Evans, Christopher. "Notes on the opensurf library." *University of Bristol, Tech. Rep. CSTR-09-001*, January (2009).
- [20] CAMPBELL, Colin. Kernel methods: a survey of current techniques. *Neurocomputing*, 2002, 48.1: 63-84.
- [21] Prof. Olga Veksler, CS434a541a Pattern Recognition (Available at: http://www.csd.uwo.ca/faculty/olga/Courses//CS434a_541a//Lecture11.pdf).