

Descriptor Clustering SURF for Bag of Visual Words Representation in Fingerprint Images Using K-MEANS and Hierarchical Agglomerative Clustering

Antony Eka Aditya
Universitas Dian Nuswantoro
Teknik Informatika S1
Semarang
antonyekaaditya@gmail.com

Catur Supriyanto, S.Kom, M.CS
Universitas Dian Nuswantoro
Semarang
catur.dinus@gmail.com

Abstract—Pengenalan sidik jari adalah salah satu sifat biometric yang populer digunakan untuk mengenali seseorang. Untuk mengenali identitas seseorang melalui sidik jari, perlu adanya sebuah metode yang tepat dalam melakukan identifikasi. Beberapa teknik diusulkan pada penelitian sebelumnya untuk pengenalan sidik jari. *Bag of visual word* ini memerlukan sebuah *clustering* terhadap beberapa *keypoint* yang dihasilkan dari sebuah algoritma *matching point*. Kemudian setelah *keypoint* dikelompokkan, maka *keypoint* tersebut akan digunakan untuk proses klasifikasi. Penelitian ini hanya bertujuan untuk mengetahui kinerja algoritma clustering dalam pengelompokan *keypoint*. Hierarchical Agglomerative Clustering dan K-Means dipilih untuk proses clustering tersebut. Dalam penelitian yang dilakukan algoritma K-Means mempunyai kelemahan dalam melakukan evaluasi. Sedangkan algoritma Hierarchical Agglomerative Clustering dalam melakukan evaluasi membutuhkan waktu komputasi klastering yang cukup cepat, namun hasil performa clustering *keypoint*-nya tidak cukup baik.

Keywords : Sidik jari, *Bag of visual words*, *Clustering*, *K-Means*, *Keypoint*, *Hierarchical agglomerative clustering*

I. PENDAHULUAN

Pengenalan sidik jari adalah salah satu sifat biometric yang populer digunakan dalam investigasi kriminal, sistem control akses, kartu identitas (KTP), ATM dan instalasi keamanan lainnya [1]. Dalam bidang keamanan yang digunakan sekarang, masih menggunakan keamanan konvensional seperti Password dan PIN. Namun, ada masalah dengan menggunakan keamanan yang bersifat konvensional [2]. Clustering adalah sebagai divisi dari kelompok data yang heterogen menjadi subkelompok homogen yang disebut “kelompok” [3]. Speeded Up Robust Feature (SURF) adalah ciri lokal detektor dan deskripsi yang dapat digunakan untuk tugas-tugas seperti objek pengakuan

atau 3D rekonstruksi. Hal ini sebagian terinspirasi oleh skala-invarian fitur transform [4]. *Bag of visual words (BOVW)* adalah kumpulan *keypoint* yang di ekstraksi sebagai patch gambar menonjol dan sering digunakan dalam klasifikasi data citra [5]. Dalam penelitian ini peneliti mencoba melakukan eksperimen mengevaluasi performa dari *clustering keypoint* dengan menggunakan algoritma *Hierarchical Agglomerative Clustering (HAC)*. Kemudian penulis melakukan analisa dari hasil performanya untuk dibandingkan dengan algoritma K-Means.

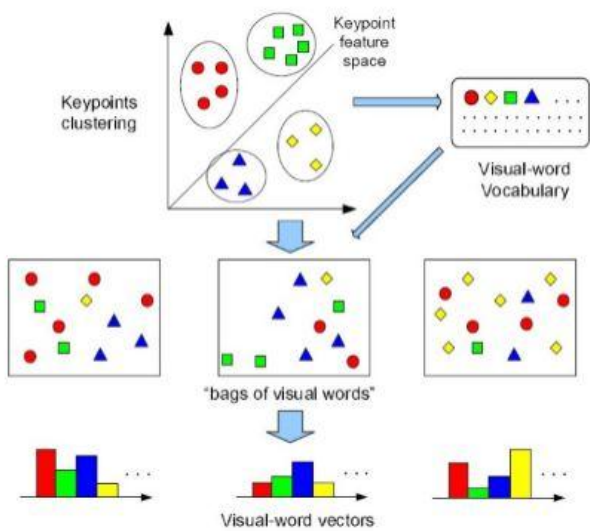
II. LANDASAN TEORI

A. Speeded Up Robust Feature (SURF)

Algoritma SURF [6] bertujuan untuk mendeteksi fitur local suatu citra dengan handal dan cepat. Algoritma ini sebagian terinspirasi oleh algoritma SIFT (*scale-invariant feature transform*), terutama pada tahap *scale space representation*. Dalam implementasinya, algoritma SURF dibagi menjadi beberapa tahapan yaitu : *Interest point detection*, *Scale space representation*, *Feature description* dan *Feature matching*.

B. Bag of Visual Word

Bag of visual words (BOVW) digunakan di bidang computer vision [7]. BOVW sering digunakan dalam pengenalan dan penyaringan pada citra. Metode BOVW mempunyai tiga tahapan [8]: 1) Ekstrak descriptor fitur local (seperti SIFT), 2) Mengkuantisasi descriptor lokal ke codebook, 3) Konversi gambar menjadi satu set kata visual (*vocabulary*). Berikut contoh prosesnya di gambarkan pada gambar 1 :



Gambar 1. Bag of Visual Words

C. K-Means

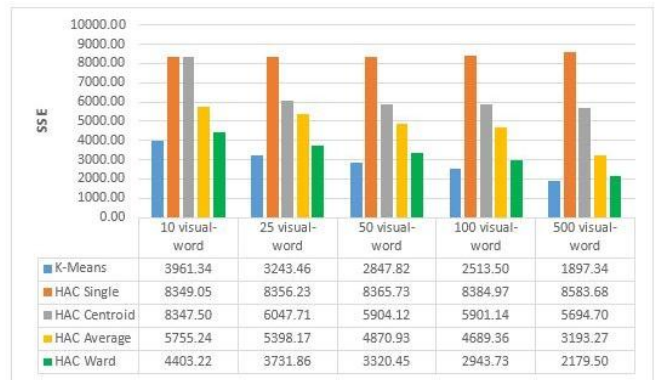
K-Means adalah salah satu algoritma klusterisasi tertua dan dikembangkan oleh JB Mac Ratu pada tahun 1967 [9]. Algoritma *K-Means* memiliki kemampuan yang baik untuk menangani banyak data yang dievaluasi. Hal ini paling cocok untuk menciptakan bentuk kurva seperti bentuk yang di inginkan untuk pendekatan yang terbaik. Algoritma *k-means* biasanya digunakan dalam data mining dan pengenalan pola dan bertujuan untuk meminimalkan indeks kinerja *cluster* [10].

D. Hierarchical Agglomerative Clustering

Algoritma SURF [6] bertujuan untuk mendeteksi fitur local suatu citra dengan handal dan cepat. Algoritma ini sebagian terinspirasi oleh algoritma SIFT (*scale-invariant feature transform*), terutama pada tahap *scale space representation*. Dalam implementasinya, algoritma SURF dibagi menjadi beberapa tahapan yaitu : *Interest point detection*, *Scale space representation*, *Feature description* dan *Feature matching*.

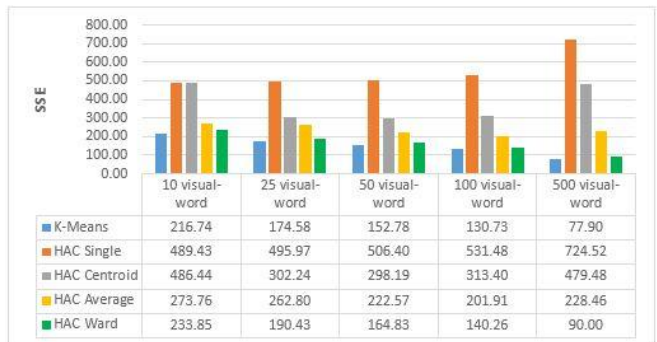
III. EKSPERIMEN

Pada eksperimen ini peneliti menggunakan spesifikasi komputer RAM 6GB dan Prosesor Intel Core i5-460M. Dan objek yang di teliti yaitu citra sidik jari yang di kelompokkan menjadi 4 dataset, yaitu : dB 1, dB 2, dB 3 dan dB 4. Namun dalam proses eksperimen yang dilakukan untuk dB 2 tidak bisa dilakukan evaluasinya, dikarenakan spesifikasi komputer yang kurang memadai. Sehingga hanya 3 Dataset yang digunakan.



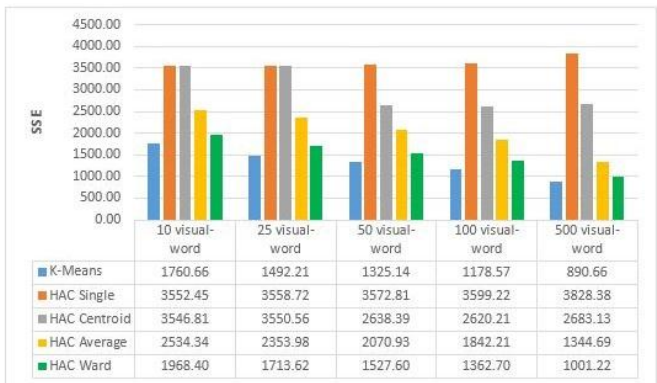
Gambar 1. SSE SURF-64 dB 1

Pada gambar 1. Algoritma K-means lebih baik dalam menentukan hasil SSE SURF-64 pada dB 1, karena nilai SSE yang lebih kecil dari pada algoritma HAC.



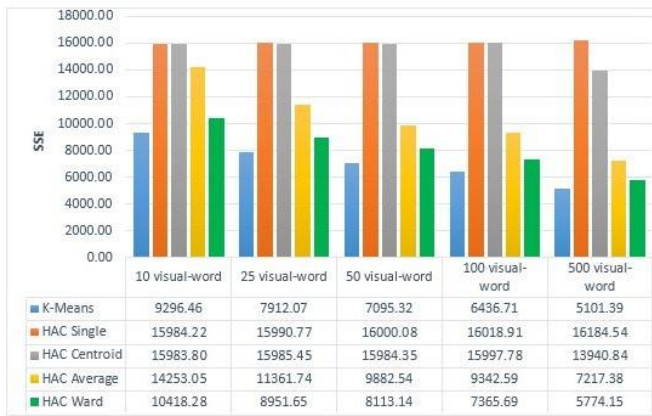
Gambar 2. SSE SURF-64 dB 3

Pada gambar 2. Memperlihatkan bahwa algoritma K-means dominan mempunyai nilai yang paling kecil dalam menentukan SSE SURF-64 pada dB 3. Sedangkan algoritma HAC mempunyai nilai yang tinggi dalam proses menentukan hasil SSE.



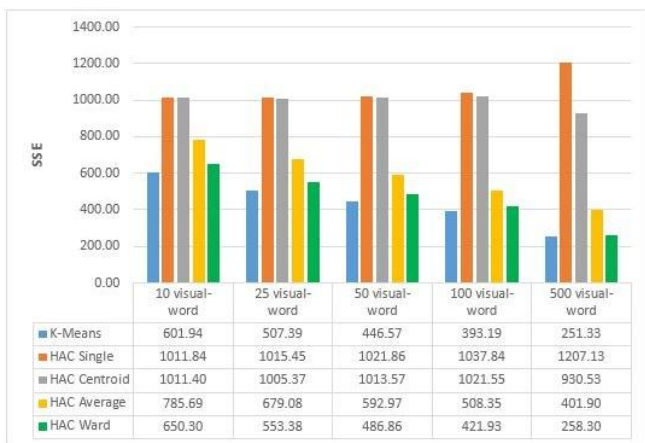
Gambar 3. SSE SURF-64 dB 4

Pada gambar 3. Menjelaskan bahwa algoritma K-means lebih baik dalam menentukan hasil SSE SURF-64 pada dB 4, Namun pada algoritma HAC Ward dalam menentukan hasil SSE memperoleh nilai yang cukup rendah hanya selisih sedikit dari algoritma K-means.



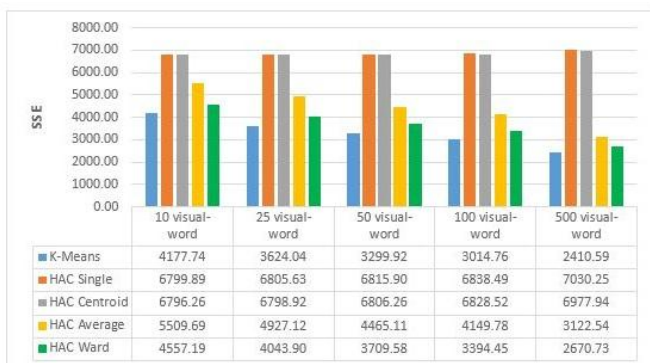
Gambar 4. SSE SURF-128 dB 1

Pada gambar 4. Algoritma K-means lebih baik dalam menentukan hasil SSE SURF-128 pada dB 1 daripada algoritma HAC.



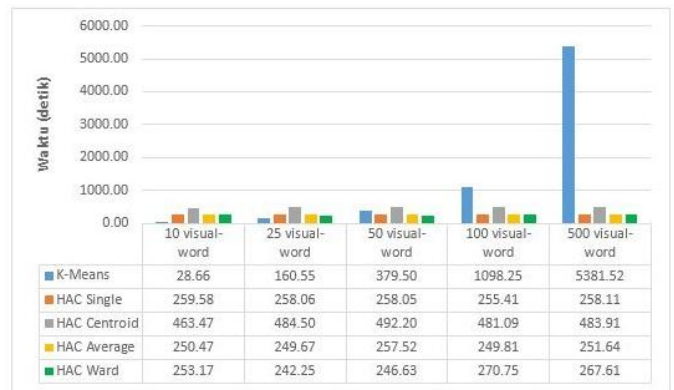
Gambar 5. SSE SURF-128 dB 3

Pada gambar 5. Algoritma K-means dan Algoritma HAC Ward hanya mempunyai nilai SSE yang terbilang selisih sedikit, Namun algoritma K-means lebih unggul dalam menentukan hasil SSE SURF-128 pada dB 3.



Gambar 6. SSE SURF-128 dB 4

Pada gambar 6. Menjelaskan bahwa algoritma K-means memperoleh nilai SSE yang cukup baik, Sedangkan Algoritma HAC dari ke empat jenisnya memperoleh nilai SSE yang cukup tinggi.



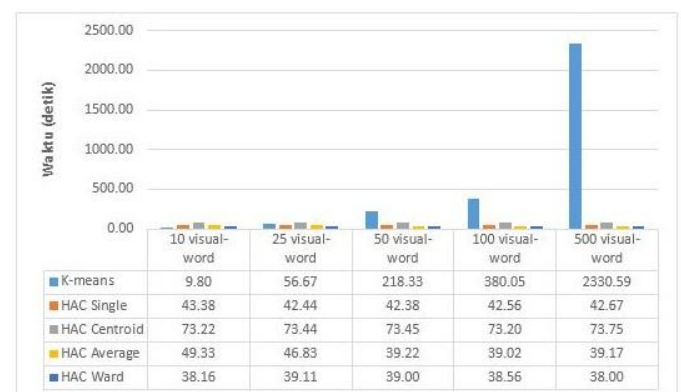
Gambar 7. Waktu Komputasi SURF SSE-64 dB 1

Pada gambar 7. Algoritma K-means pada visual-word yang ke-10 memerlukan waktu komputasi yang cukup cepat, namun bertambahnya visual-word hingga ke-500 membutuhkan waktu yang cukup lama. Sedangkan untuk algoritma HAC membutuhkan waktu komputasi yang cukup cepat.



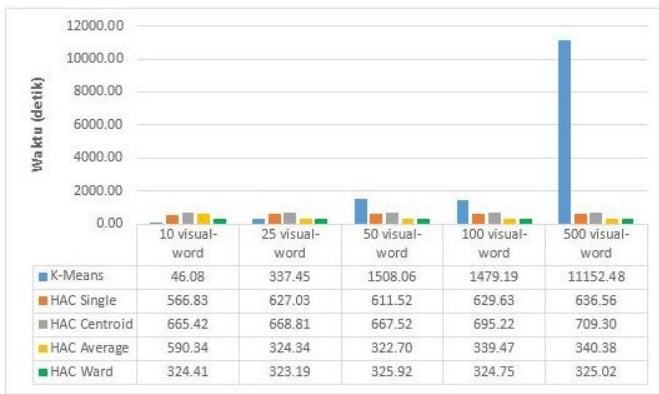
Gambar 8. Waktu Komputasi SURF SSE-64 dB 3

Pada gambar 8. Algoritma HAC Ward membutuhkan waktu komputasi yang cukup cepat dibandingkan dengan algoritma k-means yang membutuhkan waktu yang cukup lama.



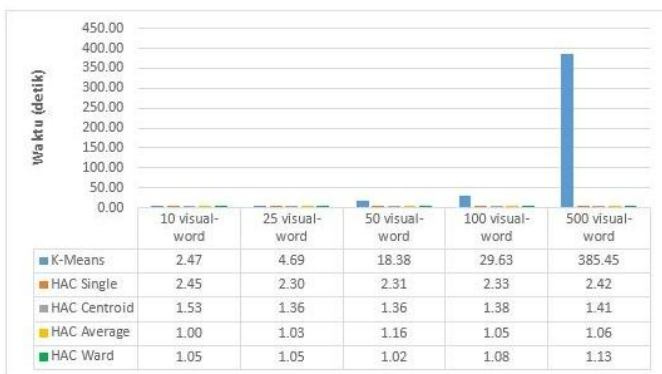
Gambar 9. Waktu Komputasi SURF SSE-64 dB 4

Pada gambar 9. Algoritma K-Means pada visual-word 10 mempunyai waktu komputasi yang paling cepat di banding algoritma HAC. Namun algoritma K-means membutuhkan waktu yang cukup lama ketika visual-wordnya bertambah hingga ke 500. Sedangkan algoritma HAC hanya membutuhkan waktu komputasi yang cepat untuk mengevaluasi visual-word ke 25, 50, 100 dan 500.



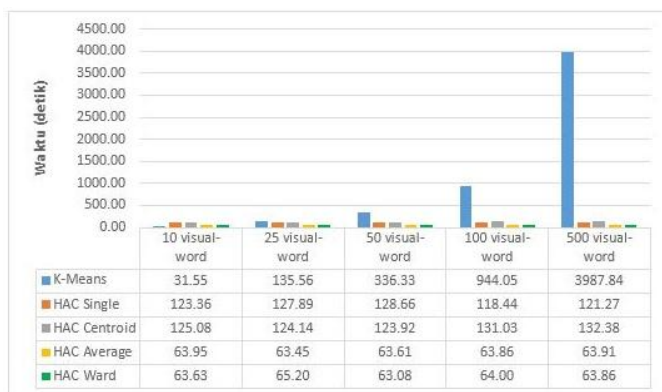
Gambar 10. Waktu Komputasi SURF SSE-128 dB 1

Pada gambar 10. Bahwa algoritma HAC Ward lebih cepat dalam melakukan evaluasi performa clustering keypoint. Sedangkan algoritma K-Means membutuhkan waktu komputasi yang lama pada visual-word 50 sampai visual-word 500.



Gambar 11. Waktu Komputasi SURF SSE-128 dB 3

Pada gambar 11. Menjelaskan bahwa algoritma HAC Ward masih lebih cepat untuk mengevaluasi performa clustering keypoint pada SURF SSE-128 di dB 3. Sedangkan algoritma K-Means masih membutuhkan waktu yang cukup lama dalam proses evaluasinya.



Gambar 12. Waktu Komputasi SURF SSE-128 dB 4

Pada gambar 12. Algoritma K-Means masih membutuhkan waktu komputasi yang cukup lama terutama pada visual-word ke 50 sampai ke 500. Sedangkan algoritma HAC Ward membutuhkan waktu komputasi yang cepat.

IV. KESIMPULAN DAN SARAN

I. Kesimpulan

Algoritma K-means jauh lebih baik dalam menentukan hasil SSE pada setiap dataset. Sedangkan algoritma HAC masih belum cukup baik untuk menentukan hasil SSE. Namun waktu komputasi yang dibutuhkan pada algoritma k-means masih cukup lama. Sedangkan waktu komputasi yang dibutuhkan algoritma HAC cukup cepat.

II. Saran

Perlu adanya pengembangan terkait menentukan hasil SSE yang terbaik menggunakan algoritma yang lain. Dan perlu adanya metode reduksi dimensi untuk mempercepat waktu komputasi. Metode yang dapat diusulkan seperti *Singular Value Decomposition (SVD)* dan *Non-negative Matrix Factorization (NMF)*

V. REFERENCES

- [1] M. U. Munir and M. Y. J. a. S. A. Khan, "A hierarchical k-means clustering based fingerprint quality classification," *Neurocomputing*, vol. 85, pp. 62-67, 2012.
- [2] Madhuri and R. Mishra, "Fingerprint Recognition using Robust Local Features," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 6, 2012.
- [3] K. Savas and K. Yildiz, "A Web Based Clustering Analysis Toolbox (WBCA) design Using MATLAB," *Procedia Social and Behavioral Sciences*, vol. 2, pp. 5276-5280, 2010.
- [4] H. Bay and T. T. a. L. V. Gool, "SURF : Speeded Up Robust Features," *ETH Zurich, Katholieke Universiteit Leuven*.
- [5] Z. L. a. X. Feng, "Near Duplicate Image Detecting Algorithm based on Bag of Visual Word Model," *Journal of Multimedia*, vol. 8, no. 5, 2013.
- [6] A. A. S. Gunawan, P. G. A and a. W. Gazali, "Pendeteksian Rambu Lalu Lintas dengan Algoritma Speeded Up Robust Features (SURF)," *Matematika Statistik*, vol. 13, no. 2, pp. 91-96, 2013.
- [7] J. Yang, Y.-G. Jiang and A. H. a. C.-W. Ngo, "Evaluating Bag-of-Visual-Words Representations in Scene Classification," *Computer Sciene Department*, vol. 9, 2007.
- [8] J. Zhang, L. Sui, L. Zhuo, Z. Li and a. Y. Yang, "An approach of bag-of-word based on visual attention model for pornographic images recognition in compressed domain," *Neurocomputing*, vol. 110, pp. 145-152, 2013.

- [9] A. Kumar, P. K. Jain and a. P. M. Pathak, "Curve Reconstruction of Digitized Surface using K-Means Algorithm," *Procedia Engineering*, vol. 69, pp. 544-549, 2014.
- [10] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Physics Procedia*, vol. 25, pp. 1104-1109, 2012.