

## **KLASIFIKASI KATEGORI CITRA DIGITAL DENGAN METODE BAG OF VISUAL WORDS**

**Mahardika Abdi Prawira Tanjung**

Program Studi Sistem Informasi, STMIK Royal Kisaran

*email:* dika.abdi@gmail.com

**Abstract:** The human eye can distinguish objects from digital images, however, computers do not have the ability as human eyes that can directly distinguish objects from digital images. Therefore the bag of visual words method was created. Bag of visual words is a method for presenting digital images based on local features. Bag of visual words illustrates how an image can be taken its characteristics, so that computers can distinguish objects on digital images. The test results show that the bag of visual words are still not maximal in classifying digital image categories, especially the chair category, which is only able to produce the most accurate accuracy of 75%. To improve the performance quality of bag of visual words in classifying digital image categories, especially the chair category, you can add an approach to determine the good number of K in clustering the visual words pattern.

**Keywords:** Bag Of Visual Words, Classification, Digital Image, Speed-Up Robust Feature, Support Vector Machine

**Abstrak:** Secara kasat mata manusia bisa membedakan objek pada citra digital, namun, komputer tidak memiliki kemampuan sebagai mata manusia yang dapat secara langsung membedakan objek pada citra digital. Maka dari itu diciptakanlah metode bag of visual words. *Bag of visual words* adalah metode untuk menyajikan citra digital berdasarkan fitur lokal. Bag of visual words menggambarkan bagaimana suatu gambar dapat diambil karakteristiknya, sehingga komputer dapat membedakan objek pada citra digital. Hasil pengujian menunjukkan bag of visual words masih belum maksimal dalam mengklasifikasi kategori citra digital khususnya kategori *chair*, yang hanya mampu menghasilkan akurasi paling akurat sebesar 75 %. Untuk meningkatkan kualitas kinerja bag of visual words dalam mengklasifikasi kategori citra digital khususnya kategori *chair*, dapat menambahkan pendekatan untuk menentukan jumlah K yang baik dalam mengkluster pola visual words.

**Kata kunci:** Bag Of Visual Words, Klasifikasi, Citra Digital, Speed-Up Robust Feature, Support Vector Machine

## PENDAHULUAN

*Bag of visual words* adalah metode untuk menyajikan gambar berdasarkan fitur lokal [1]. *Bag of visual words* menggambarkan bagaimana suatu gambar dapat diambil karakteristiknya. Setiap gambar memiliki karakteristik tertentu dalam nilai pikselnya, seperti yang ditunjukkan pada Gambar 1 dan Gambar 2 dalam bentuk gambar mobil dan objek lainnya.



Gambar 1 Citra dengan objek-objek.



Gambar 2 Citra dengan objek-objek.

Secara kasat mata manusia bisa membedakan bahwa dalam gambar itu ada mobil dan benda lain, seperti tanaman, tanah dan rumah yang terdapat dalam gambar. Namun, komputer tidak memiliki kemampuan sebagai mata manusia yang dapat secara langsung membedakan objek pada gambar. Karena

komputer hanya mengenali angka biner yaitu 0 dan 1. Untuk itu *Bag of visual words* dapat digunakan untuk mengambil nilai piksel pada gambar sebagai fitur karakteristik dalam membedakan objek dalam gambar.

Tahapan *Bag of visual words* dimulai dengan ekstraksi fitur lokal citra, ekstraksi fitur lokal dalam *Bag of visual words* dibagi menjadi 2 bagian, yaitu deteksi interest point dan deskripsi interest point [2]. Ekstraksi fitur adalah proses pemilihan informasi pada gambar yang mendingin, baik bentuk, warna, maupun sudut objek pada gambar dengan menggunakan metode tertentu [3]. Ekstraksi fitur adalah tahap pra-pemrosesan yang penting dalam klasifikasi, untuk menentukan fitur dataset yang akan diklasifikasikan. Secara teoritis, pemilihan fitur dapat meminimalkan tingkat kesalahan klasifikasi dan jumlah fitur untuk menjadi pengetahuan klasifikasi referensi [4].

Dalam ekstraksi fitur lokal ini, beberapa metode telah dikembangkan seperti menggunakan detektor [5], gambar segmentasi grid biasa dengan garis horizontal dan vertikal untuk mendapatkan area blur lokal [6], scale invariant feature transform [7], speed-up robust features [8] dan variabel fitur persegi panjang berdasarkan area blur lokal.

Dalam prosesnya skema *bag of visual words* mempunyai 4 tahapan pada klasifikasi diantaranya; menggunakan metode deteksi interest point dan ekstraksi interest point, pengklusteran fitur menggunakan k-means algorithm, pembentukan histogram generation dan klasifikasi menggunakan support vector machine sebagai klasifier (SVM) [10].

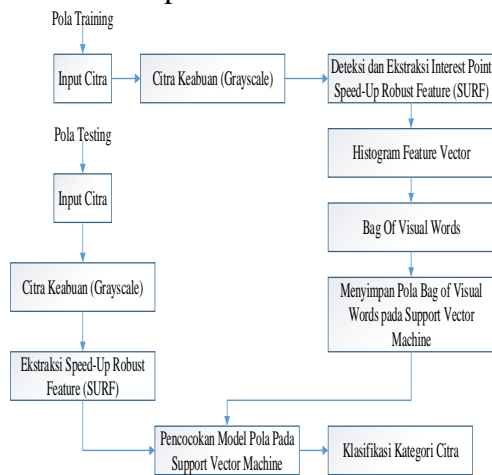
## METODE

### Datasheet

Data yang digunakan dalam penelitian ini adalah citra *imageset* Caltech101 (*brain, camera, cellphone* dan *chair*) yang berukuran 300 x 200 pixel. File citra yang digunakan untuk pelatihan maupun pengujian diambil dari situs.

### Langkah Penelitian

Langkah penelitian secara umum yang dibangun dalam penelitian ini diilustrasikan pada Gambar 3.



Gambar 3. Langkah Penelitian Secara Umum

Pada gambar 3 terdapat 2 buah proses yaitu proses training dan proses testing, pada proses training citra inputan dilakukan pre-procesing dengan grayscale dan dilanjutkan dengan mengekstraksi fitur dari citra kategori citra digital menggunakan SURF dan model pola yang telah diekstrak akan disimpan dalam Bag Of Visual Words, sedangkan pada tahapan testing, citra inputan dilakukan pre-procesing dengan grayscale dan diekstrak menggunakan SURF kemudian dilanjutkan ke tahapan pengklasifikasian model pola yang terdapat pada Bag Of Visual Words

menggunakan SVM, jika pola mirip atau mendekati pola training maka output dari klasifikasi kategori citra digital.

### 1. Speed-Up Robust Feature (SURF)

Untuk mendeteksi fitur pada citra, digunakan SURF dalam bag of visual words. SURF mengambil interest point pada citra, dimana interest point ini adalah deskripsi pada setiap bagian citra.

Penentuan interest point SURF feature menggunakan matrix hessian, dimana matrix hessian didefinisikan sebagai berikut: (Bay et al. 2006).

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

dimana  $L_{xx}(X, \sigma) = \partial^2 g(\sigma) / \partial^2 x$  adalah konvolusi dari orde kedua derivatif Gaussian dengan input citra pada point  $X = (x, y)$ , dan serupa untuk  $L_{yy}(X, \sigma)$  (Bay et al. 2006). Dengan menggunakan perkiraan orde kedua derivative Gaussian dapat dievaluasi dengan sangat cepat menggunakan citra yang diintegrasikan. Seperti yang ditunjukkan pada gambar 2.5 dengan menggunakan filter kotak 9x9 memperkirakan orde kedua derivative Gaussian dengan skala ( $\sigma=1.2$ ).

### 2. K-Means Algorithm

K-means adalah algoritma pengelompokan partial sederhana berbasis prototipe yang mencoba menemukan gugus K yang tidak tumpang tindih. Kelompok-kelompok ini diwakili oleh centroids mereka (centroid cluster biasanya adalah mean dari titik-titik di cluster itu).

Proses pengelompokan K-means adalah sebagai berikut. Pertama, centroid K awal dipilih, di mana K ditentukan oleh pengguna dan menunjukkan jumlah cluster yang diinginkan. Setiap titik dalam data kemudian ditugaskan ke

centroid terdekat, dan setiap koleksi titik yang ditugaskan ke centroid membentuk sebuah cluster. Centroid dari setiap cluster kemudian diperbarui berdasarkan titik-titik yang ditugaskan ke cluster tersebut. Proses ini diulang sampai tidak ada titik perubahan cluster (Han & Kamber, 2006). K-means dapat dinyatakan dengan fungsi objektif yang bergantung pada proximities titik data ke centroid cluster seperti berikut:

$$J = \sum_{j=1}^k \sum_{all\ i} |x_i^j - c^j|^2, \text{ in class } j$$

Dimana  $c_k$  menyatakan koordinat vektor dari  $j^{th}$  kluster dan  $\{x_i^j\}$  adalah point yang ditetapkan kepada  $j^{th}$  kluster.

### 3. Histogram Generation

Tujuan dari bag of visual words adalah mendapatkan histogram yang bisa mewakili gambar yang diberikan. Oleh karena itu, persiapan akhir kami untuk klasifikasi citra adalah histogram untuk semua gambar di dalam imageset yang digunakan. Histogram sendiri adalah alur dari frekuensi relatif dari peristiwa masing-masing nilai pixel yang diizinkan pada citra terhadap nilai-nilai itu sendiri.

### 4. Support Vector Machine (SVM)

Support vector machine (SVM) adalah algoritma yang menggunakan pemetaan nonlinier untuk mentransformasikan data pelatihan asli ke dimensi yang lebih tinggi. Dalam dimensi baru ini, ia mencari linier pemisah linier yang linier (yaitu, "batas keputusan" yang memisahkan tupel dari satu kelas dari kelas yang lain). Dengan pemetaan nonlinear yang sesuai dengan dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan oleh hyperplane. SVM menemukan hyperplane ini menggunakan vektor pendukung (tupel pelatihan "esensial")

dan margin (ditentukan oleh vektor pendukung) (Nixon & Aguado, 2008). Pada klasifikasi kategori citra digital penelitian ini, svm berfungsi sebagai classifier terhadap citra dan fitur yang telah dibentuk.

## PEMBAHASAN

### 1. Sampel Pelatihan Pola Citra

Sampel pelatihan pola kategori citra yang digunakan dalam penelitian ini berjumlah 80 citra digital yang mewakili kategori Otak, Kamera, Telpn Seluler, dan Kursi. Gambar 3 melampirkan beberapa sampel pola kategori citra digital yang digunakan sebagai pelatihan.



Gambar 3. Sampel Pelatihan Pola Kategori Citra

### 2. Penentuan Koordinat dan Deteksi Interest Point

Penentuan koordinat *interest point* dimulai dengan pembentukan grid pada citra dengan menggunakan grid step 8 yang gunanya untuk mengatur jarak koordinat x (width) dan y (height), dimana koordinat penentuan Interest Point selanjutnya akan berjarak 8 pixel.



Gambar 4. Penentuan koordinat *interest point*

Pada gambar 4 dapat dilihat hasil dari penentuan koordinat interest point, dimana perpindahan koordinat disesuaikan dengan penentuan grid stepnya, pada gambar 3.2 grid step yang digunakan 8, sehingga terlihat jelas hasil garis pembatas dari penentuan koordinat interest point. Dapat dilihat hasil dari pendeteksian interest point pada gambar 5.



Gambar 5. Hasil Deteksi Pola *Interest Point*

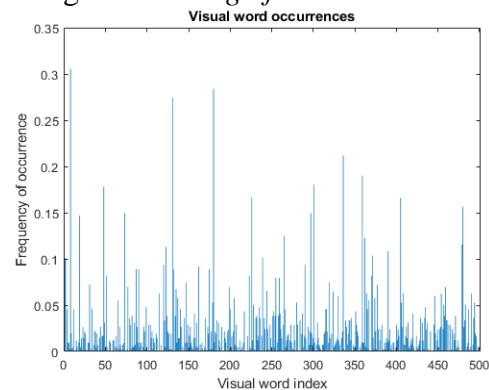
Setiap pola kategori citra digital memiliki tingkat kejenuhan masing-masing, sehingga hasil dari pendeteksian interest point pun berbeda-beda, pada gambar 4 terdapat beberapa pola hasil pendeteksian interest point yang mewakili *value* daripada kategori citra digital tersebut.

### 3. Pengujian Klasifikasi Kategori Citra Digital

Klasifikasi akan di uji cobakan dengan 5 kali pengujian pada bag of visual words. Pengujian dimulai dengan menentukan jumlah trainingset dan validation set. Pembagian citra terhadap kedua set tersebut sebesar 60 % sebagai trainingset dan 40 % sebagai validation set secara *random* atau acak.

Maka sebanyak 120 citra digital pada keseluruhan kategori citra dijadikan sebagai trainingset. Dan 80 citra lainnya dijadikan sebagai validation set. Validation set berfungsi layaknya testing set pada umumnya, untuk menguji keakuratan klasifikasi kategori citra digital terhadap 4 kategori berbeda yaitu *brain*, *camera*, *cellphone* dan *chair*.

Selanjutnya setelah pembagian trainingset dan validation set, maka dilakukan deteksi interest point dengan metode SURF, lalu menggunakan K-means clustering untuk menemukan visual vocabulary yang ditampilkan dengan *histogram generation* dan menghasilkan *bag of visual words*.



Gambar 6. Hasil *Bag Of Visual Words* Pada Pengujian Pertama

Pada tahap akhir citra pada validation set akan diuji menggunakan SVM terhadap *bag of visual words* yang telah dibentuk. Berikut ini adalah hasil



akhir pengujian klasifikasi kategori citra digital sebanyak 5 kali:

Tabel 1. Hasil Pengujian Pertama Kategori Citra Digital

Asal Mula	Diprediksikan			
	Brain	Camera	Cellphone	Chair
Brain	<b>0.85</b>	0.15	0.00	0.00
Camera	0.00	<b>0.95</b>	0.00	0.05
Cellphone	0.20	0.00	<b>0.80</b>	0.00
Chair	0.20	0.20	0.00	<b>0.60</b>

Tabel 2. Hasil Pengujian Kedua Kategori Citra Digital

Asal Mula	Diprediksikan			
	Brain	Camera	Cellphone	Chair
Brain	<b>0.90</b>	0.10	0.00	0.00
Camera	0.00	<b>0.95</b>	0.00	0.05
Cellphone	0.20	0.00	<b>1.00</b>	0.00
Chair	0.15	0.20	0.00	<b>0.65</b>

Tabel 3. Hasil Pengujian Ketiga Kategori Citra Digital

Asal Mula	Diprediksikan			
	Brain	Camera	Cellphone	Chair
Brain	<b>0.95</b>	0.15	0.00	0.05
Camera	0.00	<b>0.90</b>	0.00	0.10
Cellphone	0.05	0.00	<b>0.95</b>	0.00
Chair	0.25	0.35	0.00	<b>0.40</b>

Tabel 4. Hasil Pengujian Keempat Kategori Citra Digital

Asal Mula	Diprediksikan			
	Brain	Camera	Cellphone	Chair
Brain	<b>0.80</b>	0.05	0.00	0.15
Camera	0.00	<b>0.95</b>	0.00	0.05
Cellphone	0.15	0.00	<b>0.85</b>	0.00
Chair	0.10	0.25	0.00	<b>0.65</b>

Tabel 5. Hasil Pengujian Kelima Kategori Citra Digital

Asal Mula	Diprediksikan			
	Brain	Camera	Cellphone	Chair
Brain	<b>0.90</b>	0.05	0.00	0.00
Camera	0.00	<b>0.90</b>	0.00	0.10
Cellphone	0.05	0.00	<b>0.95</b>	0.00
Chair	0.15	0.10	0.00	<b>0.75</b>

Pada tabel 1, menunjukkan hasil klasifikasi kategori citra digital menggunakan SVM, SVM mampu mengklasifikasi dengan baik yaitu dengan tingkat akurasi 85 % pada kategori *Brain*, pada kategori *Camera* SVM mampu mengklasifikasi dengan tingkat akurasi 95 %, pada kategori *Cellphone* SVM mampu mengklasifikasi dengan tingkat akurasi 80 % dan pada kategori *Chair* SVM mampu mengklasifikasi dengan tingkat akurasi 60 %. Dan pada pengujian pertama rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital sebesar 80 %.

Sedangkan pada tabel 2, menunjukkan hasil klasifikasi kategori citra digital menggunakan SVM, SVM mampu mengklasifikasi dengan baik yaitu dengan tingkat akurasi 90 % pada kategori *Brain*, pada kategori *Camera*

SVM mampu mengklasifikasi dengan tingkat akurasi 95 %, pada kategori *Cellphone* SVM mampu mengklasifikasi dengan tingkat akurasi 100 % dan pada kategori *Chair* SVM mampu mengklasifikasi dengan tingkat akurasi 65 %. Dan pada pengujian pertama rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital sebesar 88 %.

Sedangkan pada tabel 3, menunjukkan hasil klasifikasi kategori citra digital menggunakan SVM, SVM mampu mengklasifikasi dengan baik yaitu dengan tingkat akurasi 95 % pada kategori *Brain*, pada kategori *Camera* SVM mampu mengklasifikasi dengan tingkat akurasi 90 %, pada kategori *Cellphone* SVM mampu mengklasifikasi dengan tingkat akurasi 95 % dan pada kategori *Chair* SVM mampu mengklasifikasi dengan tingkat akurasi 40 %. Dan pada pengujian pertama rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital sebesar 80 %.

Sedangkan pada tabel 4, menunjukkan hasil klasifikasi kategori citra digital menggunakan SVM, SVM mampu mengklasifikasi dengan baik yaitu dengan tingkat akurasi 80 % pada kategori *Brain*, pada kategori *Camera* SVM mampu mengklasifikasi dengan tingkat akurasi 95 %, pada kategori *Cellphone* SVM mampu mengklasifikasi dengan tingkat akurasi 85 % dan pada kategori *Chair* SVM mampu mengklasifikasi dengan tingkat akurasi 65 %. Dan pada pengujian pertama rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital sebesar 81 %.

Sedangkan pada tabel 5, menunjukkan hasil klasifikasi kategori citra digital menggunakan SVM, SVM mampu mengklasifikasi dengan baik yaitu dengan tingkat akurasi 90 % pada kategori *Brain*, pada kategori *Camera* SVM mampu mengklasifikasi dengan

tingkat akurasi 90 %, pada kategori *Cellphone* SVM mampu mengklasifikasi dengan tingkat akurasi 95 % dan pada kategori *Chair* SVM mampu mengklasifikasi dengan tingkat akurasi 75 %. Dan pada pengujian pertama rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital sebesar 88 %.

Maka rata-rata akurasi SVM dalam mengklasifikasi kategori citra digital dalam 5 kali percobaan adalah 83,4 %.

## SIMPULAN

Hasil penelitian menunjukkan bag of visual words masih belum maksimal dalam mengklasifikasi kategori citra digital khususnya kategori *chair*, yang hanya mampu menghasilkan akurasi paling akurat sebesar 75 %.

Untuk meningkatkan kualitas kinerja bag of visual words dalam mengklasifikasi kategori citra digital khususnya kategori *chair*, dapat menambahkan pendekatan untuk menentukan jumlah K yang baik dalam mengkluster pola visual words. Penambahan pendekatan tersebut tentunya akan mempengaruhi kecepatan komputasi pada pengenalan pola.

## DAFTAR PUSTAKA

- [1] Farhangi, Mohammad Mehdi, Soryani, Mohsen, & Fathy, Mahmood. 2013. Informative Visual Words Construction to Improve Bag of Words Image Representation. *IET Image Process* 8(5):310-318.
- [2] Sivic, J. Zisserman. 2003. Video google: A Text Retrieval Approach to Object Matching in

- Videos. *International Conference on Computer Vision* 2:1470-1477.
- [3] Nixon, Mark S. & Aguado, Alberto S. 2008. Feature Extraction and Image Processing. Elsevier: London.
- [4] Han, Jiawei & Kamber, Micheline. 2006. Data Mining: Concepts and Techniques. 2<sup>nd</sup> Edition. Morgan Kaufmann: Amsterdam.
- [5] Harris, C. & Stephens, M. 1988. A Combined Corner and Edge Detector. *Proceedings International Conference Alvey Vision*. 147-151.
- [6] Salomon, Chris & Breckon, Toby. 2011. Fundamentals of Digital Image Processing. Wiley-Blackwell: Oxford.
- [7] Lowe, K.D. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal Computer Vision* 60(2):91-110.
- [8] Bay, Herbert, Tuytelaars, Tinne & Van Gool, Luc. 2006. Surf: Speeded Up Robust Features. 9<sup>th</sup> *European Conference on Computer Vision*. 110(3):404-417.
- [9] Khrisnamoorthy, R. & Punidha, R. 2012. An Orthogonal Polynomials Transform-Based Variable Block Size Adaptive Vector Quantization for Color Image Coding. *IET Image Process* 6(6):635-646.
- [10] Gonzalez, R. & Woods, R. 2008. Digital Image Processing. 3<sup>rd</sup> Edition. Prentice Hall: New Jersey.