

# PENGUKURAN KEMIRIPAN SEMANTIK CITRA DENGAN 'BAG OF WORDS' LOKAL FRAGMENT SEMANTIK CITRA

## IMAGE SEMANTIC SIMILARITY MEASUREMENT USING 'BAG OF WORDS' LOCAL IMAGE FRAGMENT

Yeni Herdiyeni<sup>1</sup>, Rahmat Widyanto<sup>2</sup>, Belawati H. Widjaja<sup>3</sup>

<sup>1</sup>Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor, Kampus Darmaga, Wing 20 Level V Bogor, Jawa Barat, Indonesia. E-mail: yeni\_herdi@ipb.ac.id

<sup>2,3</sup>Fakultas Ilmu Komputer, Universitas Indonesia, Kampus Depok, Jawa Barat, Indonesia.

E-mail: <sup>2</sup>widyanto@cs.ui.ac.id, <sup>3</sup>bela@cs.ui.ac.id

Received: 15 October 2009, Accepted for publication: 21 December 2009

### INTISARI

Penelitian ini mengusulkan sebuah metode baru pengukuran kemiripan citra menggunakan visual citra "bag-of-words" fragmen lokal semantik citra. Pada metode ini, citra direpresentasikan dengan sederetan fragmen lokal citra. Fragmen lokal citra adalah bagian citra yang memiliki informasi penting dan secara konsisten muncul bersama pada citra. Ekstraksi fragmen lokal citra menggunakan pembelajaran secara *online*. Kemiripan semantik citra diukur dengan 1) mencari korespondensi antara dua pasang deretan fragmen citra; dan 2) menghitung kemiripan antara fragmen. Uji coba dilakukan dengan menggunakan *database* citra *benchmark* dengan 2,268 citra. Hasil percobaan menunjukkan bahwa metode yang diusulkan dapat meningkatkan rata-rata *precision* sebesar 74%. Hasil percobaan menunjukkan bahwa metodologi yang diusulkan dapat digunakan untuk mengukur kemiripan semantik citra dan sesuai dengan persepsi visual manusia.

**Kata kunci:** *bag-of-words*, fragmen lokal, kemiripan citra, semantik citra

### ABSTRACT

*This research proposes a new approach for measuring semantic image similarities between visual image using 'bag-of-words' semantic local image fragments. In this approach, image representation is represented using an ordered list of local image fragment. Local image fragment is a part from images which gives a significant information and appears consistently within images. Local image fragment was extracted using online learning. Image semantic similarity is measured by 1) solving a correspondence problem between two ordered sets of feature; and 2) Calculating similarities between matched features and dissimilarities between unmatched features. The methodology has already been tested on several benchmark image databases containing 2,268 images. The experimental results show that the proposed method can increase precision by as much as 74% in average. The proposed method is promising that could be used to measure semantic similarity adequately to human visual similar perception.*

**Keywords:** *bag-of-words, local fragment, image similarity, image semantic*

## 1. PENDAHULUAN

Dua hal penting dalam pengukuran kemiripan citra adalah representasi citra dan teknik pengukuran kemiripan citra. Representasi citra yang baik adalah representasi citra yang mampu merepresentasi semantik citra dan tidak sensitif terhadap perubahan cahaya, geometris. Sedangkan teknik pengukuran citra yang baik adalah yang mampu mengukur kemiripan citra sesuai dengan persepsi manusia.

Ada tiga teknik representasi citra, yaitu global (*global-based*), regional (*region-based*), dan

lokal (*local-based*)<sup>[1]</sup>. Pada pendekatan global, citra direpresentasikan dengan seluruh bagian citra. Histogram merupakan salah satu teknik representasi citra global. Pendekatan ini sangat sensitif terhadap perubahan cahaya dan geometris, karena pada pendekatan ini mengabaikan informasi spasial citra. Pada pendekatan regional, citra direpresentasikan dengan beberapa regional citra hasil segmentasi. Pada pendekatan ini, dikenal istilah *Region of Interest* (ROI). Kelemahan pendekatan ini adalah segmentasi yang kurang baik sehingga ROI yang dihasilkan tidak merepresentasikan citra yang sebenarnya. Oleh karena itu,



pada pendekatan lokal, citra direpresentasikan dengan bagian-bagian tertentu pada citra tanpa melakukan segmentasi.

*Scale-Invariant Feature Transform* (SIFT) merupakan salah satu teknik untuk merepresentasikan citra dengan menggunakan pendekatan lokal<sup>[2]</sup>. Gambar 1 menunjukkan titik utama lokal deskriptor SIFT.

Deskriptor SIFT juga digunakan untuk membentuk fragmen atau '*visual words*' pada temu kembali citra (Gambar 2)<sup>[3]</sup>. Fragmen citra diperoleh berdasarkan titik utama (*key points*) citra dari SIFT. Setelah diperoleh fragmen, selanjutnya dilakukan perhitungan untuk mengetahui banyaknya jumlah fragmen pada setiap citra. Jumlah fragmen tersebut direpresentasikan dengan vektor frekuensi fragmen (Gambar 3).

Setelah citra direpresentasikan dengan vektor frekuensi fragmen, selanjutnya pengukuran kemiripan citra dilakukan dengan menghitung jarak vektor. Deskriptor SIFT dapat merepresentasikan fitur dengan baik karena tidak sensitif terhadap perubahan skala, rotasi, translasi, dan perubahan pencahayaan<sup>[3]</sup>. Namun, SIFT membutuhkan waktu komputasi yang tinggi dan tidak tepat digunakan untuk ekstraksi citra secara *real time*<sup>[4]</sup>.

Penelitian ini menyajikan algoritma untuk mengukur kemiripan citra dengan menggunakan '*bag-of-words*' lokal semantik fragmen. Lokal semantik fragmen diperoleh melalui proses pembelajaran dengan metode Bayes<sup>[5]</sup>. Setiap citra direpresentasikan dengan beberapa fragmen semantik citra. Selanjutnya pengukuran kemiripan citra dilakukan dengan menghitung jarak vektor.

Susunan makalah ini terdiri dari enam bab. Bab 1 menjelaskan pendahuluan, bab 2 menjelaskan ekstraksi fragmen semantik citra, bab 3 menjelaskan pengukuran kemiripan citra, bab 4 dan bab 5 menjelaskan percobaan dan hasil percobaan. Bab 6 menjelaskan kesimpulan dan pemanfaatan algoritma untuk sistem temu kembali citra.

## 2. EKSTRAKSI FRAGMENT SEMANTIK CITRA

Ekstraksi fragmen semantik citra dilakukan melalui pembelajaran dengan metode Bayes<sup>[5]</sup>. Pada pembelajaran diperlukan dua kelompok

citra, yaitu citra kelas dan citra bukan kelas. Fragmen citra dideteksi dengan menggunakan *Normalized Cross Correlation* (NCC). Teknik ini banyak digunakan untuk pencocokan *template*. Teknik *cross-correlation* dimotivasi oleh pengukuran jarak Euclid<sup>[6]</sup> dengan formula:

$$d^2_{f,t}(u,v) = \sum_{x,y} [f(x,y) - t(x-u,y-v)]^2 \quad (1)$$

dengan  $f$  adalah citra,  $t$  adalah *template* (potongan citra atau fragmen),  $x$  dan  $y$  adalah posisi piksel dan  $u$  dan  $v$  adalah posisi *template*  $t$ .

Jika nilai korelasi antara fragmen dan citra pada bagian tertentu melebihi nilai ambang batas (*threshold*) tertentu yang telah ditetapkan, maka artinya pada citra tersebut terdeteksi mengandung fragmen. Pada penelitian ini, penentuan nilai *threshold* berdasarkan nilai *Minimum Cross Entropy* (MCE)<sup>[7]</sup>. Metode MCE menggunakan teori *Kullback-Leibler* dalam menghitung jarak  $D$  antara dua distribusi peluang. Jika diasumsikan  $P = \{p_1, p_2, \dots, p_n\}$  dan  $Q = \{q_1, q_2, \dots, q_n\}$  adalah dua buah distribusi peluang maka *cross entropy* didefinisikan dengan:

$$D(P, Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (2)$$

Pada metode Bayes, data pengamatan (data pelatihan) dan parameter model dinamakan dengan variabel peubah acak<sup>[5]</sup>. Data pelatihan dinyatakan dalam bentuk vektor  $x$ ,  $X_n = \{x_1, x_2, \dots, x_n\}$  dengan  $n$  adalah jumlah data dan parameter model dinyatakan  $\theta = (\theta^{(1)}, \dots, \theta^{(d)})^T$  dengan  $d$  adalah dimensi vektor. Dengan menggunakan fungsi kepekatan peluang (*probability density function*) dan distribusi gabungan  $p(x, \theta, X_n)$  dapat diketahui parameter model  $p(x|\theta)$  sebagai berikut,

$$p(x, \theta | X_n) = p(x|\theta, X_n)p(\theta | X_n) \quad (3)$$

dan kepekatan peluang bersyarat (*posterior*) dari parameter  $\theta$  dinyatakan dengan teori Bayes,

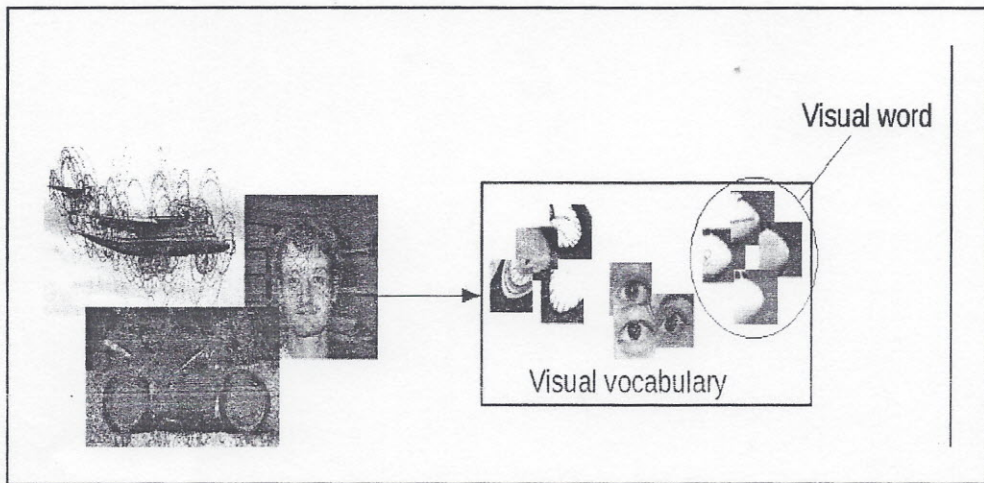
$$p(\theta | x, X_n) = \frac{p(x|\theta, X_n)p(\theta | X_n)}{p(x | X_n)} \quad (4)$$

dengan  $p(\theta|x, X_n)$  adalah *posterior*,  $p(x|\theta, X_n)$  adalah *likelihood*,  $p(\theta|X_n)$  merupakan *prior*, dan  $p(x|X_n)$  merupakan distribusi marginal/kejadian.

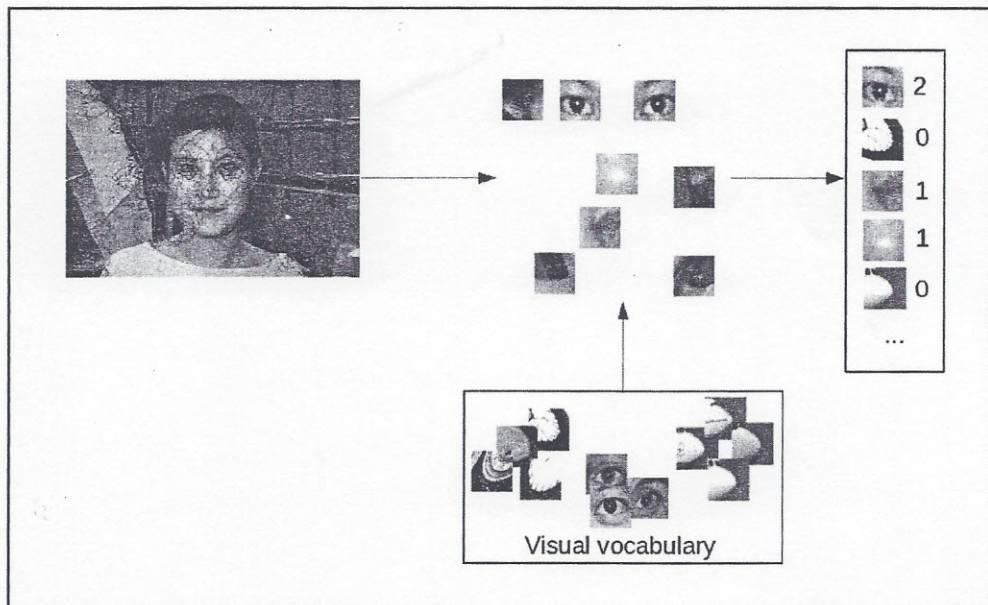




Gambar 1. Lokal deskriptor SIFT



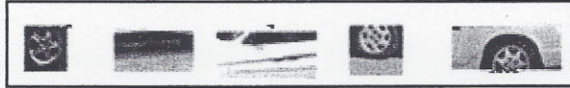
Gambar 2. Pembentukan fragmen dengan deskriptor SIFT



Gambar 3. Representasi citra dengan vektor frekuensi fragmen

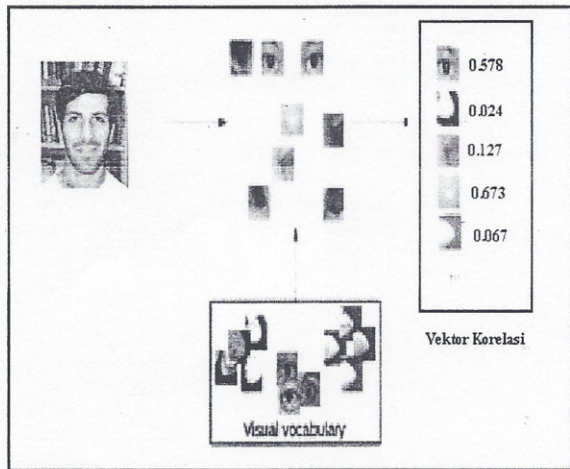


Setelah melalui tahap pembelajaran, akan diperoleh hasil seleksi beberapa potongan bagian citra yang mampu merepresentasikan semantik citra. Gambar 4 menunjukkan contoh beberapa bagian hasil deteksi semantik citra pada kelas mobil.



Gambar 4. Contoh beberapa hasil deteksi bagian semantik objek pada kelas mobil

Setelah fragmen terpilih untuk setiap kelas citra melalui proses pembelajaran, tahap selanjutnya adalah merepresentasikan setiap citra database menggunakan semua fragmen terpilih tersebut. Semua fragmen dideteksi pada setiap citra dengan menggunakan *Normalized Cross Correlation* (NCC). Setelah itu, setiap citra direpresentasikan dengan vektor korelasi fragmen (Gambar 5).



Gambar 5. Representasi citra dengan vektor korelasi fragmen

### 3. PENGUKURAN KEMIRIPAN CITRA

Pada penelitian ini, pengukuran kemiripan dilakukan menggunakan metode pengukuran jarak vektor dengan nilai vektor merepresentasikan nilai korelasi fragmen citra. Berikut ini adalah formula untuk mengukur jarak vektor menggunakan *cosine similarity*<sup>[8]</sup>:

$$\text{sim}(\vec{I}, \vec{Q}) = \frac{I^t Q}{\|I\| \|Q\|} = \frac{\sum_{i=1}^n I_i \times Q_i}{\sqrt{\sum_{i=1}^n Q_i^2} \times \sqrt{\sum_{i=1}^n I_i^2}} \quad (5)$$

dengan  $I_i$  merupakan karakteristik ke- $i$  dari sebuah citra di basis data, sedangkan  $Q_i$  merupakan karakteristik ke- $i$  dari citra kueri.

Semakin dekat nilai *cosine similarity* ke nilai 1 (satu) maka semakin mirip citra tersebut dengan citra kueri.

### 4. EVALUASI HASIL PENGUKURAN KEMIRIPAN CITRA

Pada tahap evaluasi dilakukan penilaian kinerja sistem dengan melakukan pengukuran *recall* dan *precision* untuk menentukan tingkat keefektifan hasil pengukuran kemiripan citra. *Recall* menyatakan proporsi materi relevan yang ditemukembali. Sementara itu, *precision* menyatakan proporsi materi yang ditemukembali yang relevan<sup>[9]</sup>.

$$\text{Recall} = \frac{|Ra|}{|R|} \quad \text{Precision} = \frac{|Ra|}{|A|} \quad (6)$$

dengan  $Ra$  adalah jumlah citra relevan yang ditemukembali,  $R$  adalah jumlah citra relevan yang ada di basis data, dan  $A$  adalah jumlah seluruh citra yang ditemukembali.

Untuk mengevaluasi relevansi citra kueri dengan hasil kueri juga dilakukan pengukuran *Receiver Operator Characteristic* (ROC)<sup>[10]</sup>. Pengukuran ROC dilakukan untuk mengetahui berapa jumlah kelas positif yang berhasil diklasifikasikan sejalan dengan jumlah kelas negatif yang salah diklasifikasikan. Pada penelitian ini, hasil kueri yang sesuai dengan citra kueri dinyatakan dengan kelas positif dan hasil kueri yang tidak sesuai dengan citra kueri disebut dengan kelas negatif. Tabel 1 menunjukkan matrik kontingensi untuk pengukuran ROC.

Tabel 1. Matrik kontingensi

	Aktual Positif	Aktual Negatif
Prediksi Positif	TP	FP
Prediksi Negatif	FN	TN

Berdasarkan matrik kontingensi tersebut, maka nilai *recall*, *precision*, *True Positif Rate* (TPR) dan *False Positif Rate* (FPR) dapat dinyatakan sebagai berikut:



$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

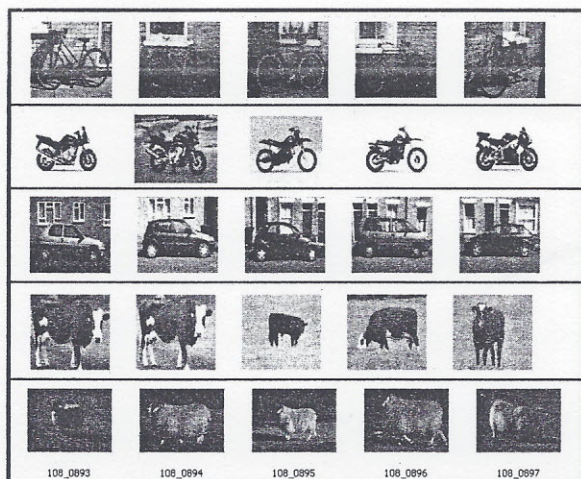
$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{TruePositifRate (TPR)} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{FalsePositifRate (FPR)} = \frac{FP}{FP + TN} \quad (10)$$

## 5. PERCOBAAN DAN HASIL PERCOBAAN

Database citra yang digunakan pada penelitian ini adalah database dari Microsoft Research Cambridge Object Recognition Image Database (MSRCORID). Pada penelitian ini ada lima kelas objek yang digunakan, yaitu kelas kuda (171 citra), mobil (423 citra), sepeda (195 citra), motor (832 citra), dan domba (148 citra). Data citra uji coba yang digunakan memiliki keragaman geometris, posisi, pencahayaan dan latar belakang citra. Hal ini dimaksudkan untuk menguji kemampuan metode yang digunakan dalam mengukur kemiripan semantik citra. Beberapa contoh citra yang digunakan pada penelitian ditunjukkan pada Gambar 6.



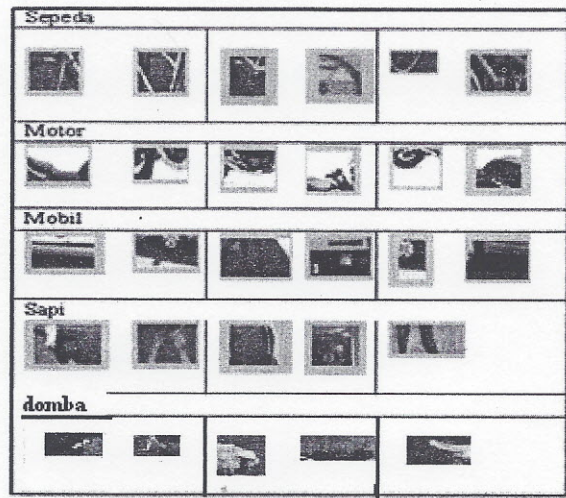
Gambar 6. Contoh citra database MSRCORID

### 5.1 EKSTRAKSI FRAGMENT SEMANTIK CITRA

Berdasarkan hasil percobaan, beberapa fragmen semantik citra ditunjukkan pada Gambar 7.

Setelah semua fragmen semantik terpilih untuk setiap kelas, tahap berikutnya adalah mendeteksi kemunculan setiap fragmen tersebut pada setiap citra dalam database. Setelah dilakukan

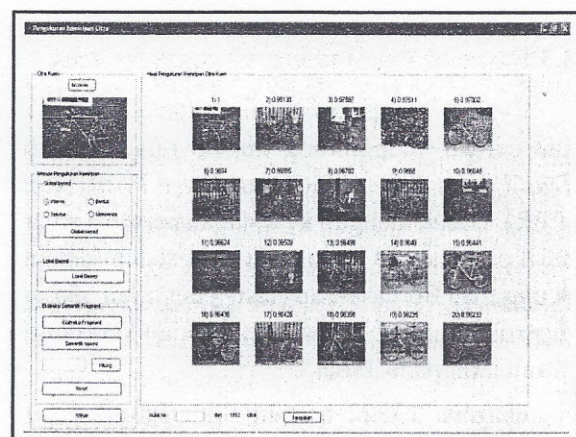
pendeteksian fragmen, maka pada akhirnya setiap citra direpresentasikan dalam bentuk vektor korelasi.



Gambar 7. Beberapa fragmen semantik citra hasil ekstraksi

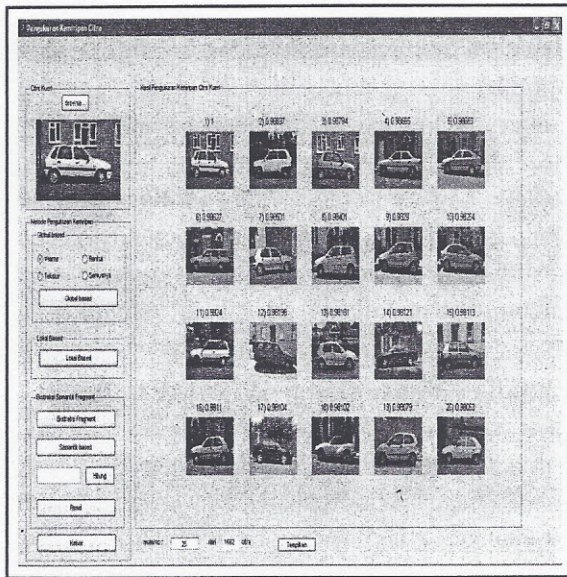
### 5.2 PENGUKURAN KEMIRIPAN CITRA

Pada penelitian ini pengukuran kemiripan citra dilakukan dengan menghitung jarak vektor menggunakan persamaan kosinus. Gambar 8 menunjukkan hasil pengukuran kemiripan citra untuk kelas sepeda. Hasil pengukuran kemiripan citra untuk kelas mobil diperlihatkan pada Gambar 9. Hasil pengukuran kemiripan citra untuk kelas motor ditunjukkan pada Gambar 10. Hasil pengukuran kemiripan citra untuk kelas kuda dan hasil pengukuran kemiripan citra untuk kelas domba masing-masing diperlihatkan pada Gambar 11 dan Gambar 12.



Gambar 8. Hasil pengukuran kemiripan citra untuk kelas sepeda

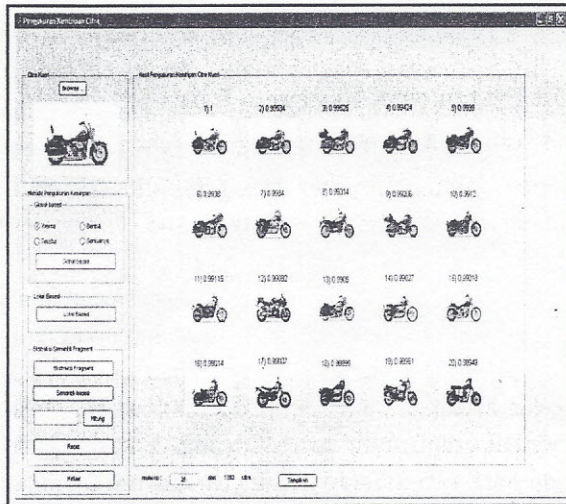




Gambar 9. Hasil pengukuran kemiripan citra untuk kelas mobil



Gambar 11. Hasil pengukuran kemiripan citra untuk kelas kuda



Gambar 10. Hasil pengukuran kemiripan citra untuk kelas motor



Gambar 12. Hasil pengukuran untuk kemiripan kelas domba

### 5.3 EVALUASI PENGUKURAN KEMIRIPAN CITRA

Evaluasi pengukuran kemiripan citra dilakukan dengan menghitung tingkat rata-rata presisi, *True Positif Rate* (TPR), dan *False Positif Rate* (FPR). Untuk menguji kesesuaian persepsi antara hasil pengukuran kemiripan citra sistem dengan pengguna. Berdasarkan hasil pengujian sistem, diperoleh tingkat rata-rata untuk setiap kelas citra, ditunjukkan pada Tabel 2.

Gambar 13–22 menunjukkan grafik *recall precision* dan kurva ROC untuk setiap kelas. Hasil percobaan menunjukkan bahwa metode semantik lokal fragmen memiliki rata-rata akurasi presisi

sebesar 0,74. Berdasarkan hasil tersebut, maka metode ini sudah cukup baik mengukur kemiripan citra berdasarkan semantik citra.

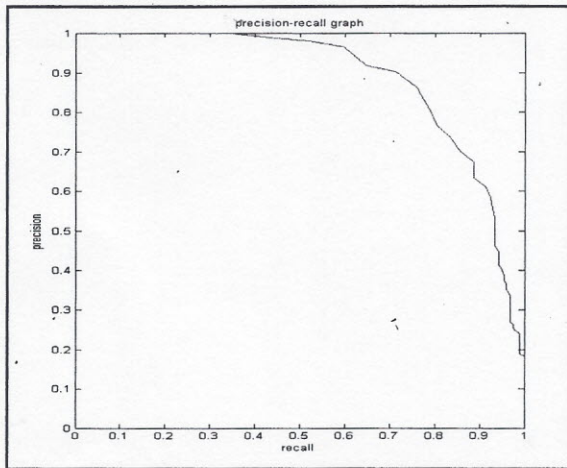
Tabel 2. Rataan presisi untuk setiap kelas citra

Kelas Citra	Presisi	TPR	FPR
Sepeda	0,81	0,81	0,29
Mobil	0,75	0,74	0,36
Motor	0,86	0,85	0,15
Kuda	0,64	0,63	0,37
Domba	0,63	0,64	0,36

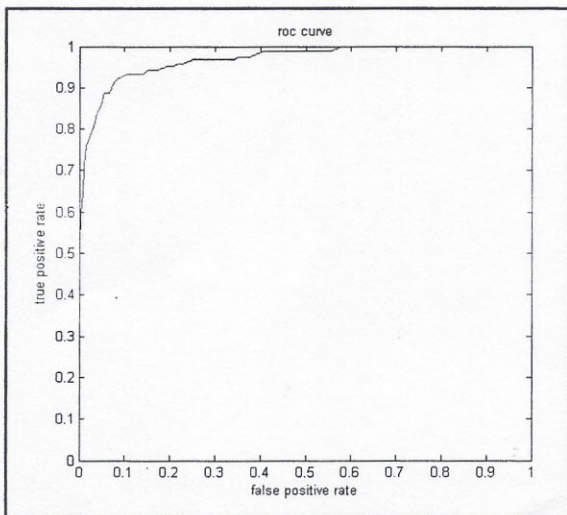
Gambar 13–14 menunjukkan grafik *recall precision* untuk citra sepeda. Untuk kelas sepeda, rata-rata akurasi presisi mencapai 0,81. Hal ini juga



ditunjukkan pada grafik ROC yang menyatakan bahwa jumlah kelas positif, yaitu hasil kueri yang relevan (TPR) mencapai 0,81 sedangkan hasil kueri yang tidak relevan (FPR) mencapai 0,29.



**Gambar 13.** Grafik *Recall Precision* untuk kelas sepeda

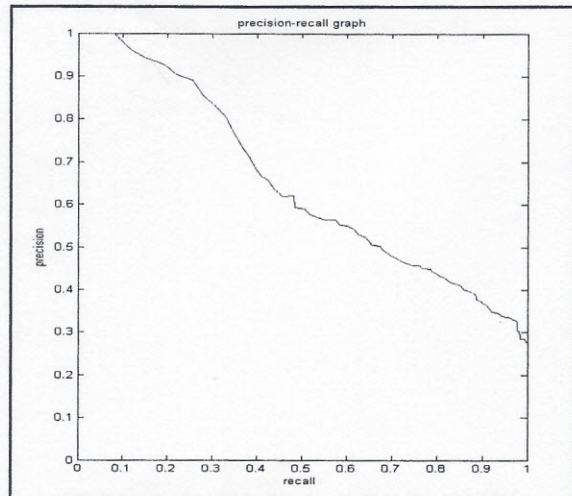


**Gambar 14.** Grafik ROC kelas sepeda

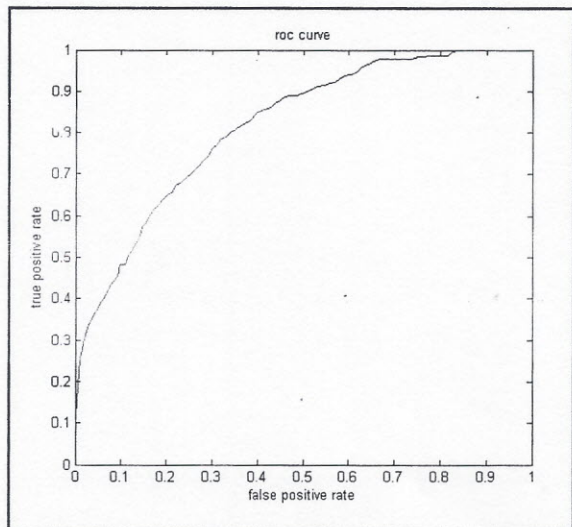
Gambar 15–16 menunjukkan grafik *recall precision* untuk citra mobil. Untuk kelas mobil, rata-rata akurasi presisi mencapai 0,75. Hal ini juga ditunjukkan pada grafik ROC yang menyatakan bahwa jumlah kelas positif, yaitu hasil kueri yang relevan (TPR) mencapai 0,74, sedangkan hasil kueri yang tidak relevan (FPR) mencapai 0,36.

Gambar 17–18 menunjukkan grafik *recall precision* untuk citra motor. Untuk kelas motor, memiliki rata-rata akurasi presisi yang paling tinggi, yaitu mencapai 0,86. Hal ini juga ditunjukkan pada grafik ROC yang menyatakan bahwa jumlah kelas positif, yaitu hasil kueri yang relevan (TPR)

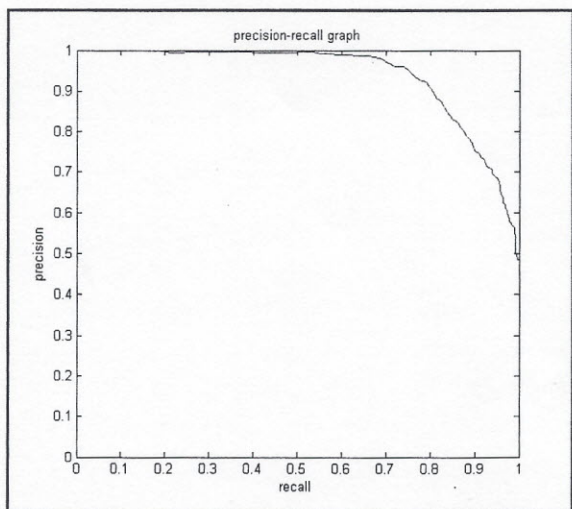
mencapai 0,85, sedangkan hasil kueri yang tidak relevan (FPR) mencapai 0,15.



**Gambar 15.** Grafik *Recall Precision* untuk kelas mobil

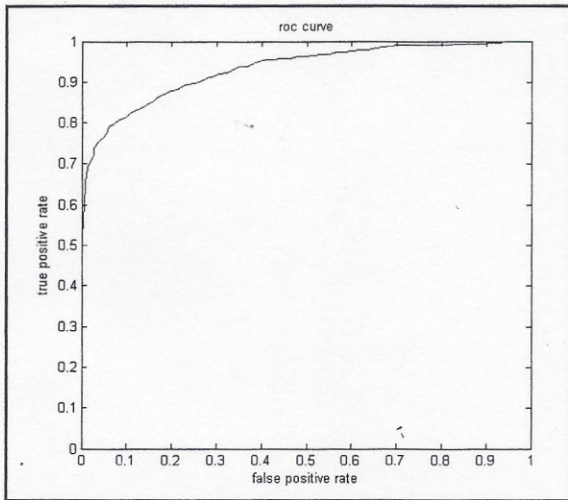


**Gambar 16.** Grafik ROC kelas mobil

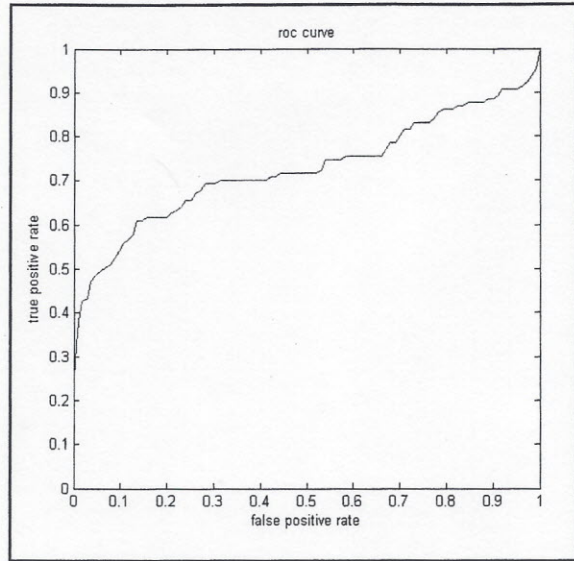


**Gambar 17.** Grafik *Recall Precision* untuk kelas motor





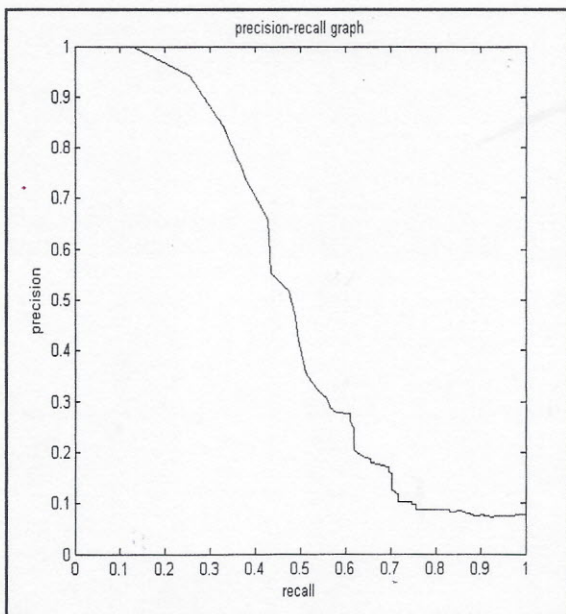
Gambar 18. Grafik ROC kelas motor



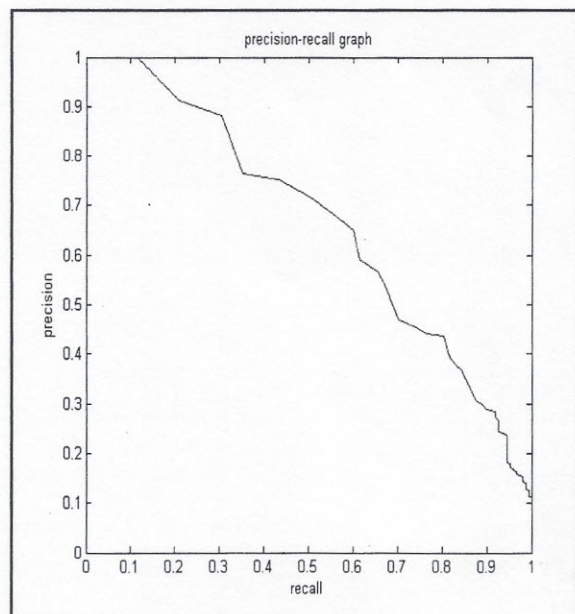
Gambar 20. Grafik ROC kelas kuda

Gambar 19–20 menunjukkan grafik *recall precision* untuk citra kuda. Untuk kelas kuda, memiliki rataan akurasi presisi mencapai 0,63. Hal ini juga ditunjukkan pada grafik ROC yang menyatakan bahwa jumlah kelas positif, yaitu hasil kueri yang relevan (TPR) mencapai 0,63, sedangkan hasil kueri yang tidak relevan (FPR) mencapai 0,47. Hasil analisis menunjukkan citra kuda yang ada di dalam *database* memiliki variasi geometris yang paling besar sehingga hasil akurasi-nya menurun dibandingkan kelas lainnya.

Gambar 21–22 menunjukkan grafik *recall precision* untuk citra domba. Untuk kelas domba, memiliki rataan akurasi presisi mencapai 0,64. Hal ini juga ditunjukkan pada grafik ROC yang menyatakan bahwa jumlah kelas positif, yaitu hasil kueri yang relevan (TPR) mencapai 0,64, sedangkan hasil kueri yang tidak relevan (FPR) mencapai 0,36. Sama seperti pada kelas kuda, hasil analisis menunjukkan citra yang ada di dalam *database* memiliki variasi geometris yang paling besar sehingga hasil akurasi-nya menurun dibandingkan kelas lainnya.

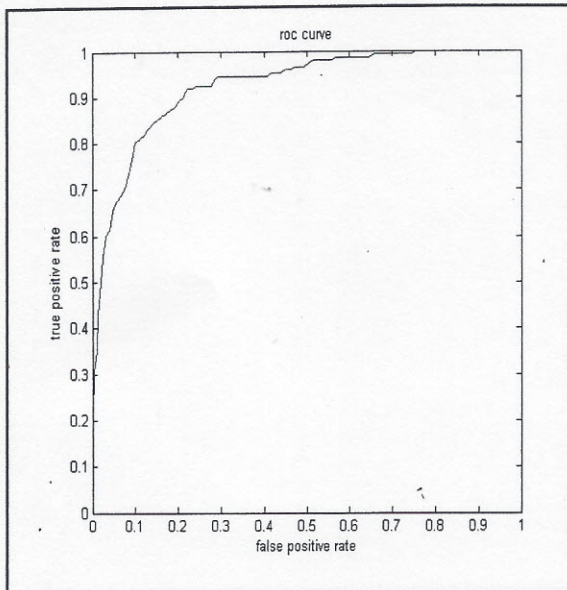


Gambar 19. Grafik *recall precision* untuk kelas kuda



Gambar 21 Grafik *recall precision* untuk kelas domba





Gambar 22. Grafik ROC untuk kelas Domba

## 6. KESIMPULAN

Penelitian ini mengusulkan metode pengukuran kemiripan semantik menggunakan ‘*bag-of-words*’ lokal fragmen semantik citra. Pada metode ini sebuah citra direpresentasikan dengan sejumlah lokal fragmen citra. Lokal fragmen citra adalah bagian dari citra yang mengandung informasi semantik citra. Lokal fragmen citra diperoleh melalui proses pembelajaran secara *online* menggunakan teknik Bayes.

Pengukuran kemiripan semantik citra terdiri dari dua tahap, yaitu: 1) Pengukuran kemiripan lokal fragmen citra dengan menggunakan *Normalized Cross Correlation* (NCC); dan 2) Pengukuran kemiripan citra menggunakan pengukuran jarak vektor kosinus. Berdasarkan hasil percobaan, metode pengukuran kemiripan citra

menggunakan ‘*bag-of-words*’ lokal fragmen semantik citra mampu meningkatkan rata-rata akurasi presisi sebesar 75%. Ini menunjukkan metode yang dikembangkan dapat mengukur kemiripan citra sesuai dengan persepsi manusia.

## 7. DAFTAR PUSTAKA

- [1] Datta, R., et al. 2007. *Image Retrieval: Ideas, Influences, and Trend of the New Age*. ACM Transaction on Computing Survey.
- [2] Lowe. 2004. *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision.
- [3] Tirilly, P., et al. 2009. *A Review of Weighting Schemes for Bag of Visual Words Image Retrieval*. Publications Internes de l’IRISA.
- [4] Jing, F., Li, M., et al. 2004a. *An efficient and effective region-based image retrieval framework*. IEEE Trans. Image Processing 13, 5, 699–709.
- [5] Fei-Fei, L., et al. 2004. *Learning Generative Visual Model from few training examples: An incremental Bayesian approach tested on 101 object categories*. In IEEE CVPR Workshop of Generative Model Based Vision (WGMBV).
- [6] Lewis, J.P. 1995. Fast Normalized Cross Correlation. *Visual Interface*, p.120-123.
- [7] Pal N.R. 1996. On Minimum Cross Entropy Thresholding. *Pattern Recognition*, vol. 29, pp. 575–580.
- [8] Smeulders, A.W.M., et al. 2000. *Content-based Image Retrieval at The End of The Early Years*. IEEE PAMI, 22(12), 1349–1380
- [9] Baeza-Yates R & Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. New York: Addison Wesley.
- [10] Davis, Jesse, dan Goadrich, Mark. 2006. *Proceeding of 34<sup>th</sup> International Conference on Machine Learning*.