PENGEMBANGAN SISTEM TEMU KEMBALI CITRA DENGAN MULTIMODAL DATA MENGGUNAKAN MICROSTRUCTURE DESCRIPTOR DAN PLSA

Choiru Za'in, Nanik Suciati, Chastine Fatichah
Institut Teknologi Sepuluh Nopember
choiruzain@gmail.com, nanik@if.its.ac.id, chastine@cs.its.ac.id

Abstrak. *Content Based Image Retrieval (CBIR) adalah sistem yang bertujuan untuk mencari sejumlah citra yang relevan berdasarkan data visual citra. Dalam perkembangannya, CBIR kemudian melibatkan pula data tekstual pada citra (multimodal data) dengan mencari korelasi antara data visual dan tekstual yang dapat dihitung menggunakan algoritma Probabilistic Latent Semantic Analysis (PLSA). Penelitian ini menggabungkan algoritma ekstraksi fitur Microstructure Descriptor (MSD) dan PLSA untuk membangun sistem CBIR multimodal data yang memiliki waktu komputasi lebih cepat. Sebagai pembanding, sistem PLSA-MSD yang dibangun dibandingkan dengan PLSA-SIFT yang digunakan pada penelitian yang telah ada. Hasil uji coba menunjukkan bahwa kombinasi PLSA-MSD 300% lebih cepat daripada PLSA-SIFT.*

Kata Kunci: *PLSA, CBIR, MSD, SIFT, Auto anotasi*

Perkembangan teknologi digital dan dukungan perkembangan kapasitas media penyimpanan dewasa ini mengakibatkan peningkatan jumlah file digital secara pesat. Peningkatan jumlah file digital terutama citra dan teks mendorong pencarian citra tertentu yang relevan dengan permintaan (*query*). Problem tersebut mendorong para peneliti untuk mengembangkan metode penemuan kembali citra yang relevan dengan efektif dan efisien. Salah satu metode yang digunakan dalam temu kembali citra adalah *Content Based Image Retrieval* (CBIR).

CBIR pada awalnya adalah sebuah proses membandingkan data visual dari citra contoh terhadap sekumpulan citra dalam suatu basis data. Namun, temu kembali citra menggunakan data visual saja terbukti tidak efisien [1]. Hal ini diakibatkan oleh pemahaman komputer terhadap konsep citra jauh lebih rendah dibandingkan dengan pemahaman manusia terhadap citra, meskipun pada tingkat pengenalan citra yang sederhana. Kelemahan komputer dalam mengenali dan memahami citra ini salah satunya disebabkan oleh keterbatasan deskripsi data visual serta masih terbatasnya metode pembelajaran komputer terhadap data visual yang ada

Berbagai riset telah dilakukan untuk meningkatkan efektifitas dan efisiensi hasil temu kembali CBIR. Usaha yang dilakukan antara lain dengan melibatkan jenis data lain yaitu data tekstual dan/atau pengembangan data visual untuk meningkatkan pengenalan citra. Penggunaan data tekstual dalam pencarian citra ini terbukti lebih efektif dalam mencari citra yang relevan [2]. Metode ini mengorganisasi teks atau deskripsi dari citra yang bersangkutan berdasarkan persamaan kata. Akan tetapi, hasil pencarian citra dengan menggunakan data teks saja lebih buruk daripada hasil pencarian menggunakan data visual saja apabila deskripsi teks bersifat global (bukan entitas, objek, manusia, atau tempat) [2]. Dalam kasus ini, jenis data teks tidak mampu menginterpretasikan jenis data visual yang ada di dalam citra. Istilah ini disebut perbedaan semantik (*semantic gap)* [3].

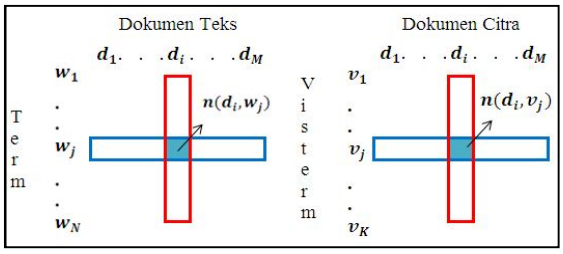
Usaha untuk mengatasi perbedaan semantik dilakukan dengan mempelajari konsep semantik dari sejumlah besar citra contoh. Salah satu pembelajaran konsep semantik dari sejumlah citra yang telah dilakukan adalah menggunakan algoritma *Probabilistic Latent Semantic Analysis* (PLSA) [4]. Pembelajaran ini dilakukan terhadap citra yang memiliki data teks (anotasi) yang berkaitan. PLSA menggunakan metode probabilistik yang menghitung probabilitas keterkaitan data visual citra terhadap data teks citra tersebut. Dengan kata lain, pembelajaran yang dilakukan oleh PLSA dapat menyimpulkan korelasi *visual term*

(visterm) citra dengan objek/kata (*tag*) yang melekat pada citra tersebut.

Sejauh ini penelitian temu kembali citra dengan multimodal data menggunakan PLSA dilakukan oleh [4] memanfaatkan fitur citra yang dibangun dengan menggunakan metode *Scale Invariant Feature Transform* (SIFT) [5]. Fitur citra ini kemudian dimanfaatkan untuk membangun visterm. SIFT memiliki keuntungan tahan terhadap transformasi, *occlusion*, iluminasi dan skala. Akan tetapi SIFT memiliki kelemahan dalam hal kecepatan komputasi. Deskriptor SIFT memiliki dimensi 128 fitur vektor pada citra *grayscale*. Dengan jumlah dimensi yang sangat besar tersebut, SIFT tidak cukup efisien untuk diterapkan pada temu kembali citra yang membutuhkan kecepatan tinggi, terutama pada ukuran basis data yang besar [6].

Dalam penelitian yang lain, *Micro-structure descriptor* (MSD) yang diusulkan oleh [6] merupakan metode ekstraksi fitur yang digunakan untuk temu kembali citra dengan lebih efisien. MSD hanya membutuhkan 72 fitur vektor untuk mendeskripsikan citra berwarna (*full color*). Formasi dari MSD, tidak hanya menginformasikan fitur tepi, akan tetapi juga warna pada fitur tepi tersebut. Dengan demikian, deskriptor MSD dapat digunakan untuk mengkombinasikan warna, tekstur, dan bentuk secara keseluruhan.

Penelitian ini menerapkan penelitian temu kembali citra dengan multimodal data menggunakan PLSA dengan menggunakan metode ekstraksi fitur MSD dengan tujuan meningkatkan efisiensi temu kembali citra. Sistem ini kemudian dibandingkan dengan PLSA dengan menggunakan metode ekstraksi fitur SIFT sebagaimana yang dilakukan oleh [4].



Gambar 1. Matriks *term*-dokumen pada *term* dan *visterm*.

Tahap awal dari penelitian ini adalah mendapatkan data visual dengan menggunakan algoritma extraksi fitur, MSD dan SIFT, dan mendapatkan anotasi citra dari data teks yang berkaitan dengan citra tersebut (*tag*). Kedua jenis data ini kemudian diproses menggunakan algoritma PLSA untuk menghitung keterkatian data-data tersebut.

Data visual dan data tekstual diproses oleh PLSA dalam bentuk matriks *term*/*visterm*-dokumen. Matriks *term*-dokumen merupakan matriks yang menunjukkan jumlah frekuensi anotasi tertentu untuk setiap dokumen teks (*term*) yang berkaitan dengan citra, sedangkan matriks *visterm* dokumen menunjukkan jumlah frekuensi *visual term* (*visterm*) tertentu untuk setiap citra. Gambaran mengenai matriks *term*/*visterm*-dokumen diilustrasikan pada Gambar 1.

Gambar 1. menunjukkan bahwa $W = \{w_1 \ldots w_N\}$ adalah sekumpulan term dimana N adalah jumlah anotasi (*term*) yang unik dari seluruh dokumen, sedangkan $V = \{v_1 \ldots v_K\}$ adalah sekumpulan visterm dimana K adalah jumlah data visual (*visterm*) yang unik dari seluruh dokumen, dan $D = \{d_1 \ldots d_M\}$ merupakan sekumpulan dokumen dimana M adalah jumlah dokumen. Setiap dokumen memiliki dua komponen, yaitu citra dan teks yang menyertai citra tersebut. Daerah pertemuan antara baris biru dan kolom merah pada dokumen teks $n(d_i, w_j)$ merupakan jumlah frekuensi *term* w_j, $1 < j < N$, pada dokumen d_i, $1 < i < M$, sedangkan $n(d_i, v_j)$ merupakan jumlah frekuensi *visterm* v_j, $1 < j < K$, pada dokumen d_i, $1 < i < M$. Matriks *term*/*visterm* dokumen *i* merupakan histogram frekuensi *term*/*visterm* dalam tiap dokumen *i*, sehingga matriks *term*/*visterm* dokumen merupakan kumpulan histogram *term*/*visterm* terhadap dokumen.

Pada awalnya, PLSA hanya digunakan untuk memproses satu jenis data (unimodal) saja [7]. Namun dalam perkembangannya PLSA dapat digunakan untuk unimodal (PLSA standar) dan multimodal data (PLSA multimodal). PLSA pada dasarnya merupakan algoritma yang mengolah matriks *term*/*visterm*-dokumen untuk mendapatkan *topik* dari distribusi *term*/*visterm* dalam dokumen. Dalam PLSA, *topik* adalah bagian dari dokumen sedangkan *term* adalah bagian dari *topik*. *Topik* dalam PLSA bersifat *latent* (tersembunyi).

Pemrosesan matriks *term*/*visterm*-dokumen oleh PLSA terdiri dari dua tahap yaitu *learning* dan inferensi. Dalam menentukan parameter *learning* dan inferensi, PLSA menggunakan

algoritma *Expectation Maximization* (EM) [7]. Parameter ini didapatkan dengan mengestimasi matriks *term*/*visterm*-dokumen menghasilkan parameter distribusi *conditional probability* dari *topik* dalam dokumen $P(z|d)$ dan *conditional probability* dari *term*/*visterm* dalam *topik* $P(x|z)$ dimana *z* merupakan variabel *topik*, *x* adalah variabel *term*/*visterm*, $x \in W$ untuk *term*, $x \in V$ untuk *visterm*, dan *d*, $d \in D$, adalah variabel dokumen. Jumlah variabel *topik* (*z*) ditentukan secara manual. Jumlah nilai distribusi variabel *z* yang didapatkan melalui algoritma EM adalah 1 (satu).

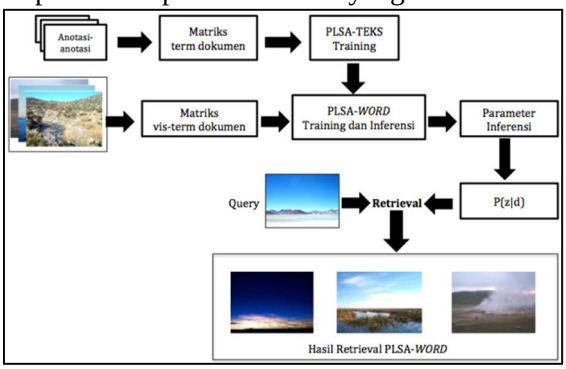
Pemrosesan PLSA standar untuk masing-masing jenis data (visual atau teks) menggunakan algoritma EM menghasilkan parameter *learning* visual maupun teks. Pada penelitian ini, variabel *learning* $P(z|d)$ dan $P(x|z)$ diberi inisial sesuai dengan proses dan jenis data yang terlibat dalam proses tersebut. sebagai contoh inisial $P(z|d)_{-visual-learning}$ dan $P(x|z)_{-visual-learning}$ untuk parameter learning dengan data visual, dan $P(z|d)_{-teks-learning}$ dan $P(x|d)_{-teks-learning}$ untuk parameter *learning* untuk data teks.

Tahap inferensi pada PLSA menggunakan data (visual atau teks) didapatkan dengan cara memproses *query* citra atau teks yang terdapat di data uji terhadap parameter *learning* sehingga menghasilkan parameter inferensi $P(z|d)_{-visual-inferensi}$ dan $P(z|d)_{-teks-inferensi}$.

Pada penelitian ini, pemrosesan PLSA unimodal data dilakukan pada data visual saja. Hasil temu kembali didapatkan dengan melakukan proses inferensi *query* citra menghasilkan $P(z|d_{query})_{-visual-inferensi}$. Kemudian, nilai inferensi *query* ini dibandingkan dengan hasil inferensi citra seluruh data uji $P(z|d)_{-visual-inferensi}$ menggunakan kedekatan jarak *cosine* (*cosine distance*). Sedangkan pemrosesan PLSA multimodal yang merupakan fokus utama dalam penelitian ini dijelaskan lebih lanjut pada Bab Metodologi.

Untuk memperjelas penamaan proses PLSA pada jenis data yang diolah, maka PLSA yang mengolah data teks disebut PLSA-TEKS, PLSA yang mengolah data visual disebut PLSA-CITRA, sedangan PLSA yang mengolah data teks dan citra disebut PLSA-*WORD*.

Sebagaimana telah didiskusikan di atas, proses ekstraksi data visual pada temu kembali citra multimodal data ini dilakukan dengan menggunakan metode ekstraksi fitur MSD dengan tujuan meningkatkan efisiensi memori dan kecepatan temu kembali sehingga dapat diaplikasikan pada basis data yang besar.



Gambar 2. Sistem Temu Kembali PLSA-*WORD* menggunakan data visual dan data teks

Selanjutnya, pada bab 1, akan dibahas metodologi sistem temu kembali PLSA yang menggunakan data visual MSD dan teks maupun PLSA yang menggunakan data visual saja. Hasil dan pembahasan akan dijelaskan pada bab 2 dan sedangkan bab 3 akan menjelaskan kesimpulan.

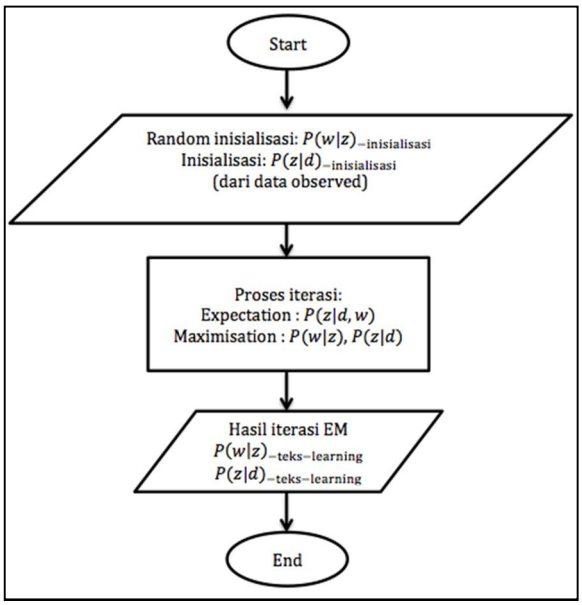
I. METODOLOGI

Sistem temu kembali multimodal data yang dibangun pada penelitian ini menggunakan *input* yang terdiri citra sebagai data visual dan anotasi sebagai data teks. *Output* yang dihasilkan adalah sekumpulan citra yang didapatkan berdasarkan *query* yang berupa citra. Desain sistem temu kembali multimodal data menggunakan PLSA ditunjukkan pada gambar 2.

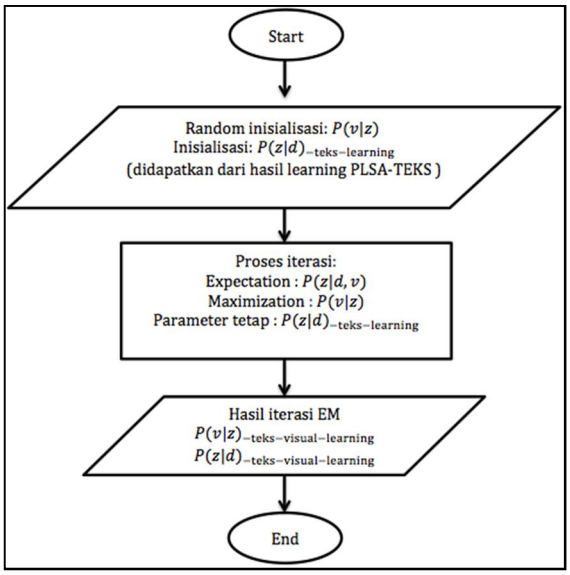
Gambar 2 terdiri dari 2 sub pemrosesan yaitu pemrosesan data teks (PLSA-TEKS) yang kemudian digabungkan dengan data visual (PLSA-*WORD*). Pada subsistem PLSA-TEKS, data teks berupa anotasi diproses menjadi matriks *term*-dokumen menghasilkan parameter *learning* teks yang dinamakan $P(z|d)_{-teks-learning}$. Diagram alir dari proses *learning* pada subsistem PLSA-TEKS ini secara terpisah dijelaskan pada gambar 3.

Hasil dari subsistem PLSA-TEKS kemudian digabungkan dengan data visual pada subsistem PLSA-*WORD*. Proses *learning* PLSA-*WORD* bertujuan untuk menghasilkan parameter *learning* visual dan teks $P(z|d)_{-teks-visual-learning}$ yang menunjukkan korelasi antara data teks dan visual. Diagram

alir proses *learning* data visual dan teks di PLSA-*WORD* diilustrasikan pada Gambar 4.



Gambar 3. Proses *learning* PLSA-TEKS pada data latih teks menghasilkan $P(z|d)_{-teks-learning}$.



Gambar 4. Proses *learning* PLSA-*WORD* pada data visual memanfaatkan informasi teks pada dataset latih menghasilkan $P(v|z)_{teks-visual-learning}$.

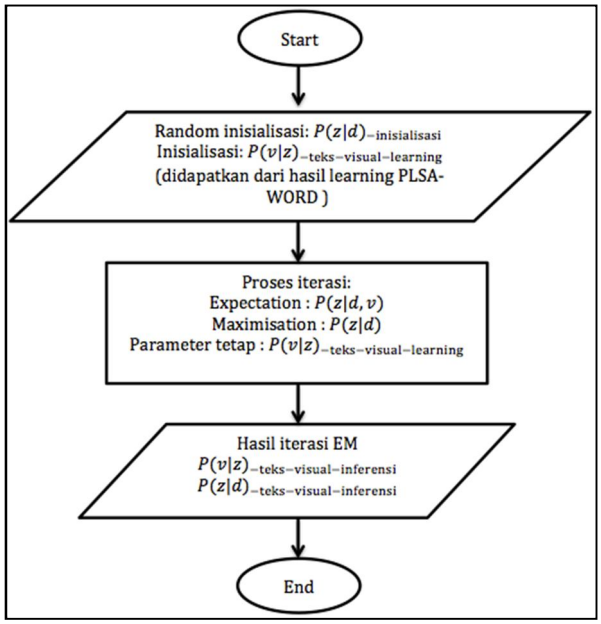
Tahap *learning* ini dilakukan pada dataset latih (*training* data). Selanjutnya, proses inferensi dilakukan dengan memproses matriks visterm dokumen citra dengan memanfaatkan parameter PLSA-*WORD* dari data latih $P(x|z)_{-teks-visual-learning}$ yang menghasilkan $P(z|d)_{-teks-visual-inferensi}$. Diagram alir proses ini diilustrasikan pada gambar 5.

Hasil temu kembali multimodal data dengan PLSA-*WORD* didapatkan dengan melakukan proses inferensi *query* citra menghasilkan $P(z|d_{query})_{-teks-visual-inferensi}$. Nilai inferensi *query* ini kemudian dibandingkan dengan hasil inferensi citra seluruh data uji $P(z|d)_{-teks-visual-inferensi}$ menggunakan kedekatan jarak *cosine*.

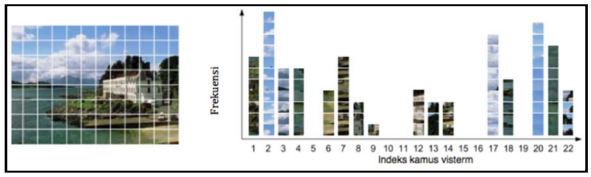
Proses pembentukan matriks visterm MSD

MSD merupakan algoritma ekstraksi fitur yang menghasilkan fitur global dimana satu citra cukup dideskripsikan oleh satu deskriptor yang memiliki dimensi 72 fitur vektor [6]. Agar dapat diproses oleh algoritma PLSA, deskriptor citra yang diproses oleh MSD harus disesuaikan agar berbentuk histogram visterm dalam dokumen. Oleh karena itu, citra perlu dipartisi menjadi *region* yang sama. Dengan demikian satu citra dapat dideskripsikan oleh banyak visterm seperti diilustrasikan pada Gambar 6.

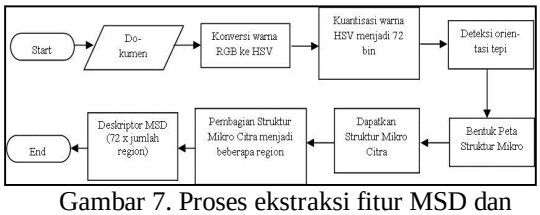
Pada penelitian ini implementasi pembagian *region* tidak dilakukan pada saat citra pertama kali diproses. Citra diekstraksi sesuai dengan urutan langkah pemrosesan MSD.



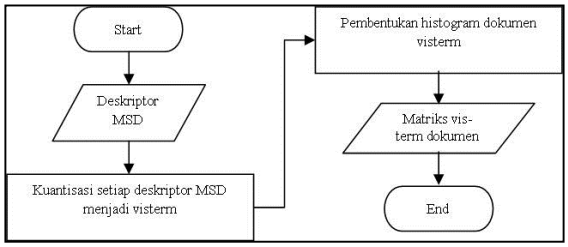
Gambar 5. Inferensi PLSA-*WORD* menghasilkan $P(z|d)_{teks-visual-inferensi}$ dari data uji yang digunakan untuk temu kembali.



Gambar 6. Representasi visterm *region* kotak dalam citra dan histogram visterm [8].



Gambar 7. Proses ekstraksi fitur MSD dan pembagian *region* pada mikro struktur citra



Gambar 8. Diagram alir pembentukan matriks visterm dokumen pada data MSD

Setelah didapatkan struktur mikro citra dalam proses ekstraksi fitur, citra dipartisi menjadi beberapa region, sehingga didapatkan deskriptor citra MSD yaitu 72 fitur vektor dikalikan dengan jumlah region dalam citra. Proses pembentukan citra dengan ekstraksi fitur MSD dalam bentuk *region* diilustrasikan pada Gambar 7.

Setelah deskriptor MSD didapatkan, deskriptor tersebut dikuantisasi menjadi sejumlah visterm. Selanjutnya setiap deskriptor diasosiasikan pada visterm hasil kuantisasi untuk mendapatkan histogram dokumen visterm. Sekumpulan histogram dokumen untuk semua citra membentuk matriks visterm dokumen. Proses pembentukan matriks visterm dokumen dari deskriptor MSD diilustrasikan pada Gambar 8.

Pengukuran kinerja

Pengukuran kinerja dievaluasi menggunakan *Average Precision* (AP) untuk setiap *query* [9]. Pengukuran ini merupakan gabungan dari pengukuran yang melibatkan nilai *Precision* dan *Recall*.

Pengukuruan kinerja temu kembali citra multimodal data dilakukan dengan menghitung presisi temu kembali yang diolah oleh PLSA-*WORD* dengan data visual yang dihasilkan oleh algoritma ekstraksi fitur MSD dan dibandingan dengan PLSA-*WORD* dengan menggunakan data visual yang dihasilkan oleh algoritma ekstraksi fitur SIFT sebagai *benchmark*.

Selain pengukuran temu kembali multimodal data, pada penelitian ini dilakukan pengukuran kinerja temu kembali unimodal data menggunakan data visual saja untuk mengetahui pengaruh data visual yang dihasilkan oleh MSD dan SIFT terhadap temu kembali tanpa menggunakan data teks.

Presisi dihitung dengan menggunakan metode *Mean Average Precision* (MAP) (Manning, Raghavan, dan Schutze, 2009), yang merupakan rata-rata dari *Average Precision* untuk setiap *query*.

Pengukuran kecepatan sistem

Pengukuran kinerja lainnya adalah waktu proses pembentukan fitur SIFT dan MSD dalam dataset, waktu proses kuantisasi visterm dari SIFT dan MSD, serta waktu proses *learning* dan inferensi pada kedua jenis fitur.

II. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini adalah SAIAPR TC-12 [12]. Dataset ini terdiri dari citra dan anotasi. Dari keseluruhan citra dalam dataset, sebagian besar citra terdiri dari anotasi yang memiliki 3 kombinasi *topik* yaitu *human*, *landscape*, dan *manmade*. Sejumlah 611 dokumen yang merupakan bagian dari keseluruhan dataset SAIAPR TC-12 dipilih sebagai dataset. Sejumlah 496 dokumen dijadikan sebagai data latih dan 115 dokumen dijadikan sebagai data uji.

Untuk menguji sistem yang dibangun, 30 *query* dipilih dari 115 dokumen uji yang didapatkan secara random. Setiap *query* yang terpilih digunakan sebagai *input* inferensi pada PLSA-CITRA yang memproses data visual saja dan *input* inferensi PLSA-*WORD* untuk multimodal data. Selain inferensi pada *query*, inferensi juga dilakukan pada seluruh data uji baik dalam sistem di PLSA-CITRA maupun PLSA-*WORD*.

Presisi hasil temu kembali unimodal data (PLSA-CITRA) dari setiap *query* dihasilkan oleh kedekatan distribusi topik dalam citra antara $P(z|d_{query})_{visual-inferensi}$ dengan $P(z|d)_{visual-inferensi}$ untuk keseluruhan data uji. Sedangkan presisi hasil temu kembali multimodal data (PLSA-*WORD*) menggunakan data visual dan teks dihasilkan berdasarkan kedekatan distribusi topik dalam citra antara $P(z|d_{query})_{-teks-visual-inferensi}$ dengan seluruh citra data uji hasil inferensi $P(z|d)_{-teks-visual-inferensi}$. Rangking kedekatan antara *query* dan dokumen di seluruh

data uji dihitung menggunakan perhitungan jarak *cosine*.

Jumlah variabel *topik* (z) dalam penelitian ini adalah 3 *topik* dalam dokumen yang merepresentasikan topik *human, landscape, dan manmade* dari anotasi dalam dokumen. Setiap variabel *topik* yang dihasilkan memiliki nilai probabilitas yang jumlah keseluruhan nilai probabilitiasnya adalah 1 (satu). Distribusi probabilitas pada *query* untuk unimodal data $P(z|d_{query})_{-visual-inferensi}$ (PLSA-CITRA) terdiri dari variabel *topik 1, topik 2, dan topik 3* dihasilkan dari inferensi PLSA-CITRA. Distribusi probabilitas pada *query* untuk multimodal data $P(z|d_{query})_{-teks-visual-inferensi}$ (PLSA-*WORD*) terdiri dari variabel *topik 1, topik 2, dan topik 3* dihasilkan dari inferensi PLSA-*WORD*.

Threshold digunakan menentukan relevansi *conditional probability* hasil inferensi *query* baik pada PLSA-CITRA maupun PLSA-*WORD*.. *Threshold* merupakan nilai batas yang digunakan menentukan relevansi dalam antara distribusi *topik* hasil inferensi citra *query* dan distribusi *topik* hasil inferensi citra data uji. Nilai *treshold* yang digunakan dalam penelitian ini adalah 0.2, 0.35, dan 0.5. Sebagai contoh, jika nilai *threshold* 0.2 diambil untuk menentukan relevansi antara *query* dan data uji, maka jika terdapat 20% distribusi *topik* hasil inferensi pada *query* dan hasil inferensi pada data uji maka *query* dan data uji itu dinilai relevan. Semakin kecil nilai *threshold* maka kemungkinan kemiripan antara citra *query* dan citra uji semakin besar dan nilai MAP yang dihasilkan besar.

Untuk setiap *query* yang dipilih akan didapatkan nilai *Average Precision* (AP) berdasarkan rangking kedekatan jarak hasil inferensi *query* dan data uji. MAP didapatkan dari rata-rata AP yang dihasilkan untuk setiap *query*. Tabel 1 menunjukkan nilai MAP.

Sebagaimana diilustrasikan pada Tabel 1, hasil uji coba menunjukkan bahwa secara umum, temu kembali multimodal data yang diproses menggunakan PLSA-*WORD* menghasilkan MAP yang lebih baik dibandingkan dengan temu kembali menggunakan data visual saja yang diproses menggunakan (PLSA-CITRA) untuk setiap *threshold* 0.2, 0.35, dan 0.5. Hasil uji coba PLSA-CITRA juga menunjukkan bahwa PLSA dengan data visual yang dihasilkan oleh SIFT menghasilkan presisi temu kembali lebih baik daripada PLSA dengan data visual yang dihasilkan MSD. Selisih MAP dari tiap *threshold* untuk PLSA-SIFT dan MSD adalah 12.8%, 16.7%, dan 18.9% untuk masing-masing *threshold* 0.2, 0.35, dan 0.5. Rata-rata selisih MAP dari PLSA-SIFT dan PLSA-MSD adalah 16.13%.

Hal ini disebabkan distribusi histogram *visterm* yang dihasilkan dari proses ekstraksi fitur dan kuantisasi elemen *visterm* SIFT lebih tersebar dibandingkan MSD. Artinya citra yang sama akan dideskripsikan oleh lebih banyak *visterm* oleh SIFT dibandingkan MSD. Histogram elemen pada data visual SIFT dan MSD diilustrasikan pada Gambar 8 dan Gambar 9.

Tabel 2 menunjukkan waktu yang dibutuhkan untuk ekstraksi fitur menggunakan SIFT dan MSD. Waktu komputasi dihitung dari proses ekstraksi fitur, kuantisasi, dan *learning* menggunakan MSD dan SIFT.

Tabel 1. MAP yang dihasilkan dari temu kembali PLSA-CITRA dan PLSA-*WORD* menggunakan data visual MSD dan SIFT menggunakan 30 *query*

Temu Kembali berbasiskan topik	MAP, dengan *threshold* yang digunakan		
	0.2	0.35	0.5
PLSA-CITRA (MSD)	0.81	0.643	0.533
PLSA-CITRA (SIFT)	0.929	0.772	0.658
PLSA-*WORD* (MSD)	1	0.968	0.929
PLSA-*WORD* (SIFT)	1	0.969	0.932

Tabel 2. Pemrosesan ekstraksi fitur MSD

No	Eks-traksi Fitur	Deskriptor yang dihasilkan	Jumlah Citra	Waktu (detik)
1	MSD	228514 deskriptor dengan 72 dimensi tiap deksriptor	611 citra ukuran 360x480	361
2	SIFT	1032578 deskriptor dengan 128 dimensi tiap deksriptor	611 citra ukuran 360x480	1421

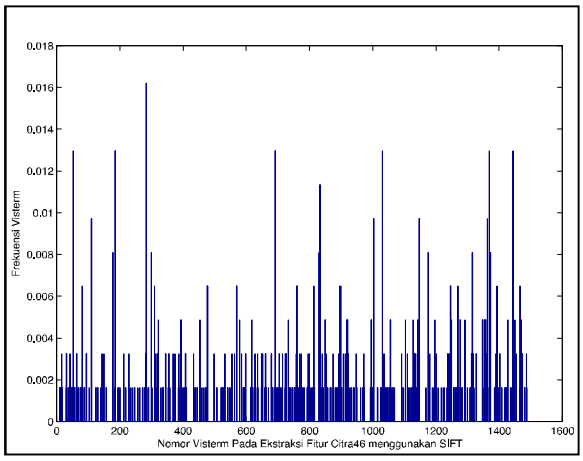
Waktu komputasi yang dilakukan dalam proses pembentukan deskriptor dengan algoritma SIFT membutuhkan waktu 3 kali lipat lebih lama dibandingkan dengan algoritma MSD. Proses pembentukan deskriptor pada 611 citra dalam dataset membutuhkan waktu 361 detik untuk ekstraksi fitur MSD dan 1421 detik untuk SIFT. Hal ini disebabkan karena jumlah deskriptor yang diekstraksi dari setiap citra menggunakan SIFT 4 kali lipat lebih banyak.

Tabel 3 menunjukkan bahwa waktu yang dibutuhkan untuk kuantisasi visterm menggunakan MSD jauh lebih efisien dibandingkan menggunakan SIFT. Waktu yang dibutuhkan MSD untuk menghasilkan 1500 visterm hanya 61.36 detik dengan 21 iterasi sedangkan pada SIFT membutuhkan 805.3 detik dengan 50 iterasi. Waktu dan jumlah iterasi yang berbeda secara signifikan ini disebabkan oleh karakteristik deskriptor yang dihasilkan oleh algoritma ekstraksi fitur. Deskriptor yang dihasilkan oleh SIFT lebih merata, sedangkan descriptor yang dihasilkan oleh MSD lebih terfokus pada satu kuantisasi warna. Dengan demikian proses kuantisasi menjadi lebih cepat.

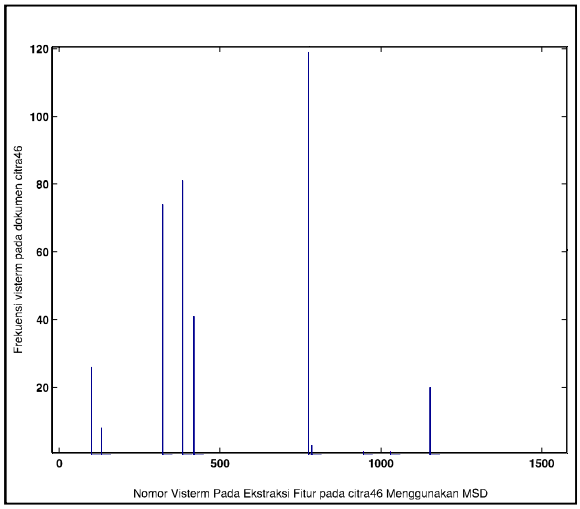
Hal ini disebabkan karena pada proses inferensi, parameter *learning* $P(z|d)_{visual-learning}$ sudah terbentuk menjadi parameter tetap untuk proses inferensi, sehingga variabel yang perlu diupdate hanyalah variabel $P(v|z)_{visual-inferensi}$. Kesimpulan pada pengukuran waktu, pemrosesan citra menggunakan PLSA-CITRA yang melibatkan fitur SIFT lebih membutuhkan waktu yang lebih lama daripada MSD.

Tabel 3. Waktu Kuantisasi pada fitur MSD dan SIFT

No	Data visual	Waktu Kuanti-sasi (detik)	Jumlah iterasi	Keterangan
1	MSD	61.36	21	228154 deskriptor dengan 72 dimensi tiap descriptor menjadi 1500 visterm
2	SIFT	805.30	50 (maks iterasi)	1032578 deskriptor dengan 72 dimensi tiap descriptor menjadi 1500 visterm.



Gambar 8. Histogram elemen pada salah satu citra uji pada data visual SIFT



Gambar 9. Histogram elemen pada salah satu citra uji pada data visual MSD

III. SIMPULAN

Berdasarkan hasil uji coba yang dilakukan, maka dapat disimpulkan bahwa waktu temu kembali multimodal data dengan menggunakan algoritma ekstraksi fitur MSD lebih cepat 300% dibandingkan waktu temu kembali multimodal data dengan menggunakan algoritma ekstraksi fitur SIFT. Pada proses *learning*, kuantisasi visterm dari deskriptor MSD hanya membutuhkan 1/13 waktu kuantisasi visterm dari deskriptor SIFT. Hal ini dikarenakan jumlah deskriptor yang dihasilkan MSD 4 kali lebih sedikit dibandingkan dengan deskriptor yang dihasilkan oleh SIFT. Akan tetapi, SIFT memiliki banyak deskriptor dan memiliki sebaran visterm yang lebih merata sehingga presisi yang dihasilkan lebih baik 3-10% dibandingkan MSD tergantung pada threshold yang digunakan.

Dengan demikian dapat disimpulkan bahwa PLSA menggunakan fitur MSD sesuai untuk data set ukuran besar, karena memiliki tingkat efisiensi waktu *learning* dan memori yang tinggi dengan tingkat presisi dengan PLSA fitur SIFT.

Hasil uji coba juga menunjukkan bahwa teks memiliki peran yang sangat besar. Hasil *learning* teks yang baik sangat mempengaruhi ketepatan akurasi tanpa tergantung pada kondisi histogram visterm (merata atau tidak merata). Hasil *learning* teks yang baik disebabkan oleh ketepatan *term* dalam mendefinisikan dokumen, selain itu *term* yang mendeskripsikan citra berupa entitas objek.

IV. DAFTAR PUSTAKA

[1] Datta, R., Joshi, D., Li, J., & Wang, J. Z.,"Image retrieval: Ideas, influences, and trends of he new age". ACM Computing Surveys (CSUR),40(2), 2008.

[2] Kherfi, M. L., Brahmi, D., & Ziou, D., "Combining visual features with semantics for a more effective image retrieval". In Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), Vol. 2:961-964, IEEE, Agustus 2004.

[3] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain., "Content-based image retrieval: the end of the early years". In IEEE Trans. on Pattern. Analysis and Machine Intelligence, 22(12):1349–1380, 2000.

[4] Lienhart, Rainer, Stefan Romberg, and Eva Hörster., "Multilayer pLSA for multimodal image retrieval." In Proceedings of the ACM International Conference on Image and Video Retrieval. ACM, 2009.

[5] Lowe, D. G., "Distinctive image features from scale-invariant keypoints". In International Journal of Computer Vision, 60(2):91-110, 2004.

[6] G-H Liu, Z-Y Li, L. Zhang, Y. Xu., “Image Retrieval Based on Micro-structured Descriptor”, Elsevier Pattern Recognition 44:2123-2133, 2011.

[7] Hofmann, T., "Unsupervised learning by probabilistic latent semantic analysis". In Machine learning, 42(1-2):177-196, 2001.

[8] Monay, F., "Learning the structure of image collections with latent aspect models", Doctoral Dissertation, Ecole Polytechnique Fédérale de Lausanne, 2007.

[9] Manning, Christoper.D, Raghavan,Prabhakar, dan Schutze. "An Introduction to Information Retrieval". Cambridge, England, Cambridge University Press, 2009.

[10] ImageCLEF, “Segmented and Annotated IAPR TC-12 Dataset”, http://imageclef.org/SIAPRdata, Agustus 2014.