

# **KAJIAN LITERATUR PERKEMBANGAN *MULTIMEDIA INFORMATION RETRIEVAL* (MIR) DAN TANTANGAN DI MASA DEPAN**

## ***A LITERATURE REVIEW ON THE DEVELOPMENT OF MULTIMEDIA INFORMATION RETRIEVAL (MIR) AND THE FUTURE CHALLENGES***

Budi Yulianto<sup>1</sup>, Rita Layona<sup>2</sup>

Fakultas Ilmu Komputer Jurusan Teknik Informatika  
Bina Nusantara University  
Jln. K.H. Syahdan No. 9, Palmerah, Jakarta Barat, 11480  
<sup>1</sup> laboratory@binus.ac.id, <sup>2</sup> rlayona@binus.edu

### **Abstrak**

*Multimedia information retrieval (MIR)* adalah proses pencarian dan pengambilan informasi (*information retrieval/IR*) dalam *content* berbentuk multimedia, seperti suara, gambar, video, dan animasi. Penelitian ini menggunakan metode kajian literatur (*literature review*) terhadap perkembangan MIR saat ini dan tantangan yang akan dihadapi di masa depan bagi para periset di bidang IR. Berbagai penelitian MIR saat ini meliputi komputasi yang berpusat pada manusia (aktor) terhadap pencarian informasi, memungkinkan mesin melakukan pembelajaran (semantik), memungkinkan mesin meminta koreksi (umpan balik), penambahan fitur atau faktor baru, penelitian pada media baru, perangkuman informasi dari *content* multimedia, pengindeksan dengan performa tinggi, dan mekanisme terhadap teknik evaluasi. Di masa yang akan datang, tantangan yang menjadi potensi penelitian MIR meliputi peran manusia yang tetap menjadi pusat (aktor) terhadap pencarian informasi, kolaborasi konten multimedia yang lebih beragam, dan penggunaan kata kunci sederhana (folksonomi).

**Kata kunci:** *multimedia information retrieval*, multimedia, komputasi, semantik, pencarian informasi

### **Abstract**

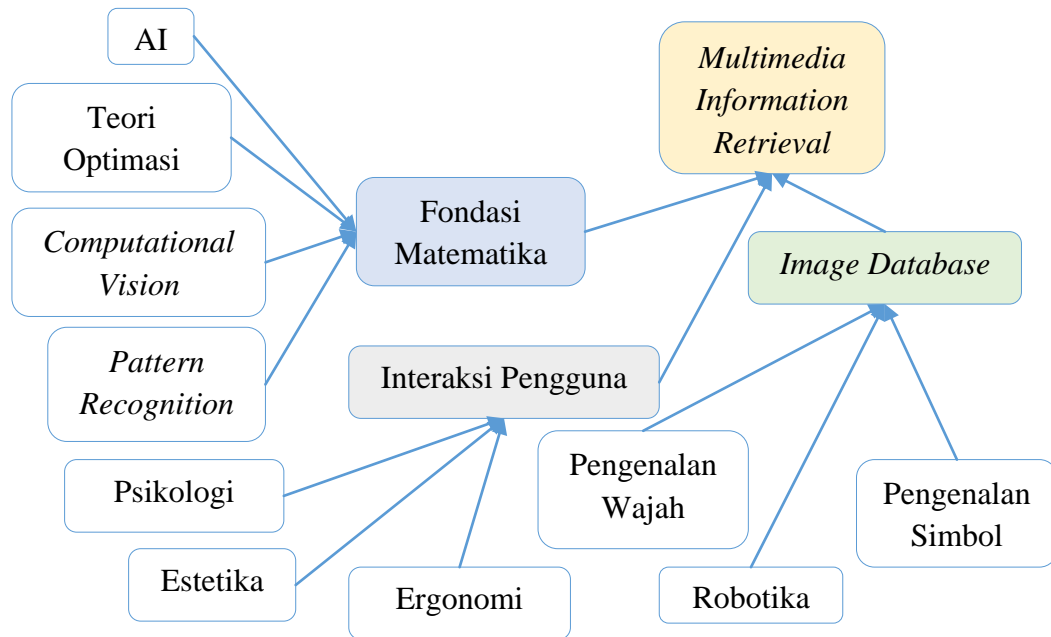
*Multimedia information retrieval (MIR)* is the process of searching and retrieving information (*information retrieval/IR*) in multimedia content, such as audio, image, video, and animation. This study uses literature review method against current MIR conditions and what challenges to be faced in the future for researchers in the field of IR. Various studies of MIR currently include human centered computation for IR, allowing machine to do the learning (semantics); allowing machine to request feedback, add new features or factors, research on new media, summarize information from multimedia content, high-performance indexing, and evaluation techniques. In the future, the potential of MIR research includes the human-centered role for information retrieval, more diverse collaborative multimedia content, and the use of simple keyword (folksonomy).

**Keywords:** *multimedia information retrieval*, multimedia, computation, semantics, information search

**Tanggal Terima Naskah** : 21 September 2015  
**Tanggal Persetujuan Naskah** : 04 Januari 2016

## 1. PENDAHULUAN

MIR merupakan pengembangan dari *traditional IR* yang hanya melakukan pencarian secara tekstual. MIR akan menjadi penting dan sangat dibutuhkan ketika tidak adanya atau kurangnya keterangan yang disisipkan dalam *content* multimedia. Namun bilamana keterangan sudah tersedia, MIR tentunya dapat digunakan untuk meningkatkan keakuratan pencarian pada *content* multimedia. MIR didukung oleh banyak bidang ilmu, seperti *artificial intelligence* (AI), *computer vision*, psikologi, estetika, robotik, dan lainnya, seperti terlihat pada Gambar 1.

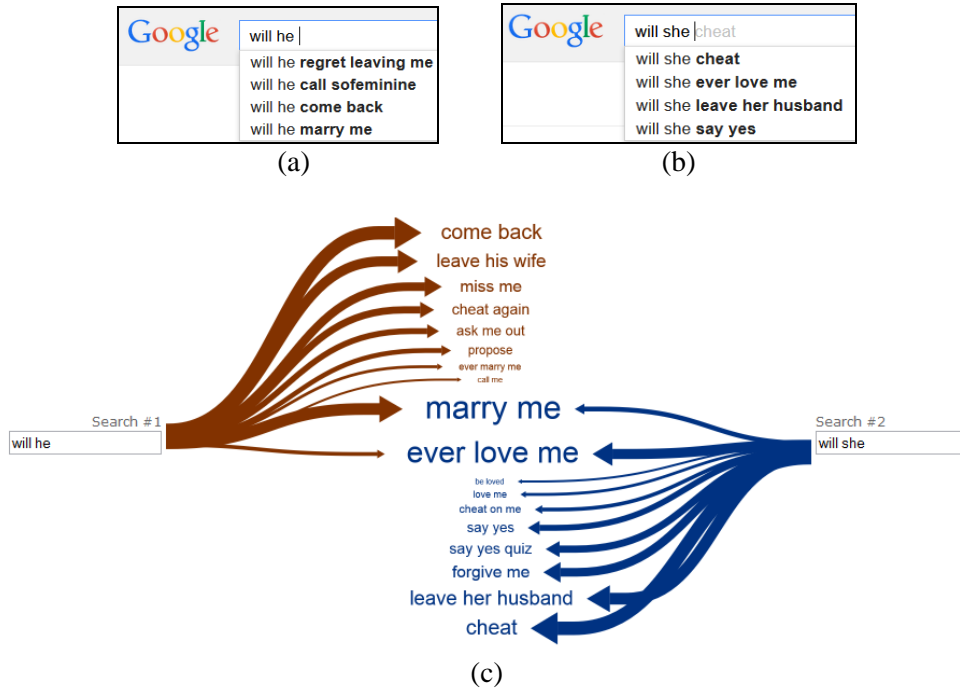


Gambar 1. Bidang ilmu pendukung MIR

Pada penelitian awal, MIR banyak didasarkan pada bidang *computer vision* [1],[2],[3]. MIR kemudian mulai memfokuskan algoritma pada pencarian kesamaan faktor atau fitur (*similarity search*) berdasarkan gambar, video, dan audio [4],[5], dan histogram warna [6]. Perkembangan selanjutnya, konsep *similarity search* ini kemudian diimplementasi pada mesin pencarian di Internet [7],[8]. Selain itu, metode *similarity search* juga diimplementasikan pada *database* perusahaan, seperti IBM DB2, atau Oracle [9],[10], yang memungkinkan MIR lebih mudah diakses oleh industri [11].

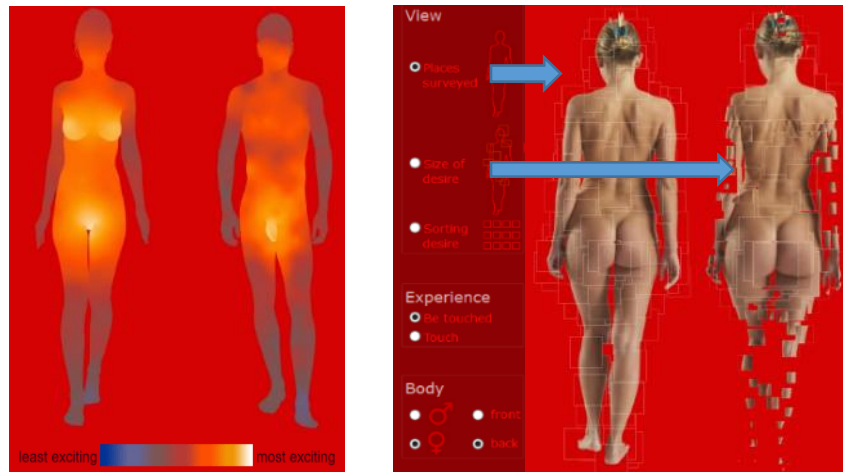
Pada abad ke 21, para peneliti menyimpulkan bahwa *feature based similarity search algorithm* sudah tidak *user-friendly*. Sistem diarahkan agar lebih *user friendly* dan dapat memberikan informasi dari *content* multimedia dengan lebih luas. Evolusi berikutnya adalah sistem perlu memahami *semantic query*, tidak hanya pada komputasi dasar tingkat rendah (*exact query*) atau pencarian tepat kata saja. Masalah ini sering disebut dengan istilah “*bridging the semantic gap*”.

Google Suggest memulai proyek *semantic query* yang memungkinkan pengguna untuk melihat apa yang orang lain cari ketika mereka melakukan pencarian pada *web*. Proyek ini mencoba memvisualisasi pikiran manusia dan memotret pertanyaan-pertanyaan yang membentuk rasa ingin tahu manusia. Riset ini diimplementasikan pada *WebSeer Project* [12].



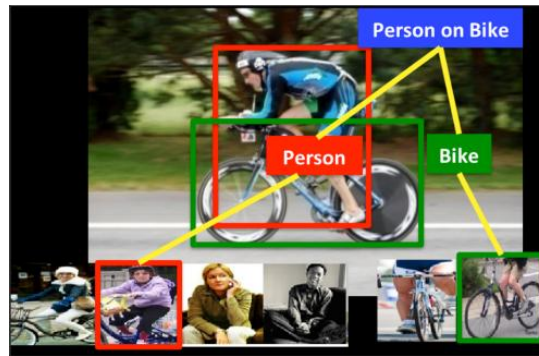
Gambar 2. (a) (b) *Google Suggest* dan (c) *WebSeer*

Selain *Google Suggest*, *Touch Project* telah melakukan proyek *semantic query* dengan menginvestigasi perspektif kolektif untuk zona sensitif seksual [13]. Sistem meminta ratusan orang untuk menentukan peringkat seberapa baik perasaan menyentuh atau disentuh oleh kekasih di berbagai titik tubuh. Gambar yang dihasilkan mengungkapkan peta keinginan sensual dengan beberapa titik fokus dan tingkat kegembiraan. Data pada *Touch Project* memetakan daerah kesenangan dan keinginan pada tubuh manusia, yang disurvei oleh ratusan pendapat orang tentang bagian-bagian tubuh (Gambar 3).



Gambar 3. Proyek visualisasi sentuhan

Perkembangan *semantic query* selanjutnya memungkinkan sistem mendeteksi *content* berbasis gambar [14]. Pengguna dapat melakukan *query* secara langsung untuk beberapa objek visual, seperti langit, pohon, air, dan lainnya menggunakan posisi objek secara spasial. Sistem menerapkan teori informasi untuk menentukan faktor terbaik dalam meminimalkan kesalahan [15].



Gambar 4. *Semantic Query* pada objek visual

Berdasarkan latar belakang perkembangan MIR saat ini, MIR di masa depan akan memiliki tantangan tersendiri. Penelitian ini mengkaji lebih dari 70 literatur (*literature review*) terhadap kondisi MIR saat ini dan merangkum apa saja tantangan yang akan dihadapi di masa depan bagi para periset di bidang IR.

## 2. MIR SAAT INI

Dua kebutuhan mendasar bagi sistem MIR adalah (1) mencari objek tertentu pada suatu media dan (2) merangkum informasi suatu koleksi media. Dalam mencari objek tertentu pada suatu media, sistem saat ini memiliki keterbatasan, seperti ketidakmampuan untuk memahami kosakata pengguna atau memahami tingkat kesesuaian hasil pencarian [16],[17],[18]. Dalam menyelesaikan dua permasalahan mendasar MIR saat ini, berbagai penelitian dilakukan, yaitu (1) komputasi yang berpusat pada manusia (aktor) terhadap pencarian informasi, (2) memungkinkan mesin melakukan pembelajaran (semantik), (3) memungkinkan mesin meminta koreksi (umpan balik), (4) penambahan fitur atau faktor baru, (5) penelitian pada media baru, (6) perangkuman informasi dari *content* multimedia, (7) pengindeksan dengan performa tinggi, dan (8) mekanisme terhadap teknik evaluasi. Dalam komputasi yang berpusat pada manusia, gagasan utama MIR adalah untuk memuaskan pengguna dan memungkinkan pengguna untuk membuat *query* dalam terminologi mereka sendiri.

### 2.1 Manusia sebagai Pusat

Sistem manusia sebagai pusat adalah dengan mempertimbangkan perilaku dan kebutuhan dari pengguna [19]. Terdapat dua studi yang menarik pada area ini. Pertama, studi terhadap pencarian gambar yang sama [20]. Pengguna diminta untuk membaca suatu artikel lokasi wisata pada sebuah *website* tentang perjalanan (*travel*) dan kemudian diminta untuk melakukan pencarian melalui gambar dan teks. Survei yang dilakukan menunjukkan bahwa 40 dari 54 responden lebih memilih untuk menggunakan teks sebagai metode pencarian lokasi daripada menggunakan gambar (Gambar 5).



atau "Manado"

Gambar 5. Pencarian berdasarkan gambar atau teks [21]

Kedua, studi tentang komputasi afektif [22],[23],[24] yang memberikan interaksi yang lebih baik dengan pengguna dengan memahami keadaan emosional pengguna. Pengenalan emosi secara otomatis dengan menggunakan model *3D wireframe* interaktif pada wajah [25],[26], atau penggalian karakter berdasarkan model yang berkaitan dengan emosi dan peristiwa pada lingkungan mereka [27],[28].



Gambar 6. Pendeteksian wajah manusia

## 2.2 Pembelajaran dan Semantik

Pembelajaran dalam MIR cukup menarik untuk menjembatani kesenjangan semantik. Literatur penelitian baru-baru ini memiliki ketertarikan yang signifikan dalam menerapkan klasifikasi dan algoritma untuk pembelajaran MIR [29],[30],[31]. Terdapat tiga studi yang menarik pada area ini. Pertama, studi mengenai deteksi objek pada latar yang kompleks. Salah satu tantangan yang paling penting dan mungkin menjadi masalah yang paling sulit dalam pemahaman semantik media adalah bagaimana mendeteksi konsep visual yang memiliki latar yang kompleks. Tantangannya adalah untuk mendeteksi semua isi semantik dalam sebuah gambar, seperti wajah, pohon, hewan, dan lainnya pada latar yang kompleks [32].



Gambar 7. Deteksi objek pada latar yang kompleks

Kedua, studi pendeteksian wajah pada latar yang kompleks. Pada pertengahan tahun 1990-an, deteksi wajah manusia pada gambar abu-abu yang memiliki latar yang kompleks telah dilakukan dengan menggunakan teori Shannon untuk meminimalisir kesalahan dalam mendeteksi wajah [33]. Riset lainnya menggunakan metode *neural networks* [34]. Selain itu, metode pendeteksian wajah berdasarkan komponen dengan menggunakan statistik juga dilakukan [35]. Pola gradien digunakan dari ekstraksi gambar video untuk mendeteksi wajah berdasarkan tingkatan kontras objek mata, hidung, dan mulut [36],[37].

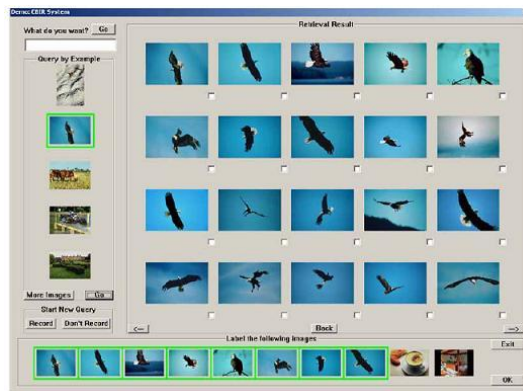


Gambar 8. Deteksi muka berdasarkan kontras objek mata, hidung, dan mulut

Ketiga, studi mengenai deteksi objek selain wajah pada latar yang kompleks. Sebuah sistem untuk mendeteksi langit, pohon, gunung, rumput dengan latar yang kompleks telah berhasil dilakukan [38]. Li dan Wang [39] menggunakan pendekatan model statistik dalam mengkonversi gambar dengan kata kunci. Rautianinen [40] menggunakan gradien dan analisis audio dalam video untuk mendeteksi objek dan melakukan pencarian semantik.

### 2.3 Umpan Balik Relevansi

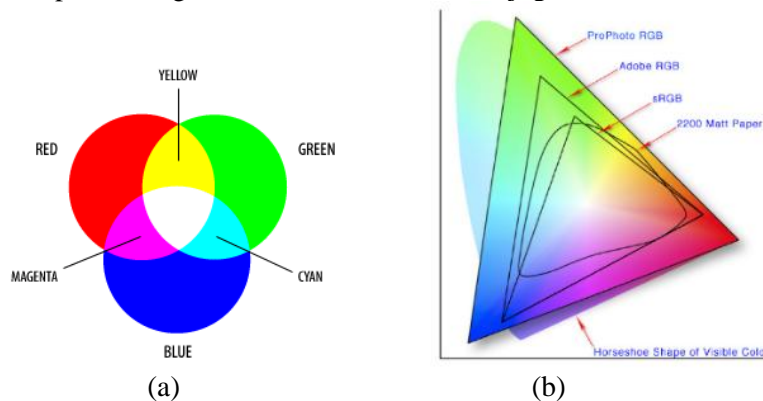
Umpan balik relevansi adalah proses interaktif yang meminta pengguna memutuskan apakah hasil pencarian sudah sesuai. Ide di balik umpan balik relevansi adalah untuk memutuskan apakah setiap hasil yang didapat sudah relevan atau tidak [41]. Chen [42] menggunakan metode SVM untuk memperbarui umpan balik. He [43] dan Yin [44] menggunakan metode *short term* dan *long term perspectives*. Metode *short term perspective* dilakukan dengan menandai tiga hasil pencarian yang tidak relevan dan 3 hasil yang relevan. Metode *long term perspective* dilakukan dengan melakukan pembaharuan pada data dari hasil metode *short term perspective* [45].



Gambar 9. Umpan balik relevansi

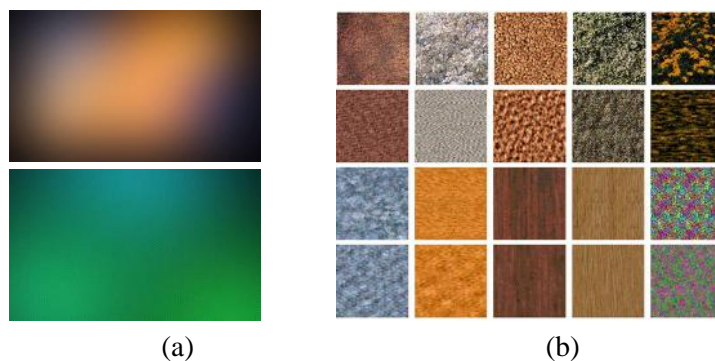
### 2.4 Faktor Baru dan Kesamaan Komponen

Penelitian tidak hanya meningkatkan algoritma pencarian, tetapi juga menciptakan faktor (fitur) baru dan langkah-langkah kesamaan komponen berdasarkan warna, tekstur, dan bentuk. Faktor warna [46],[47], seperti RGB dan *color spaces* memiliki manfaat dalam hasil pencarian berdasarkan pencahayaan [48],[49]. Pencarian dengan melakukan perbandingan kuantitatif model warna juga berhasil dilakukan [50].



Gambar 10. (a) Color model dan (b) Color space

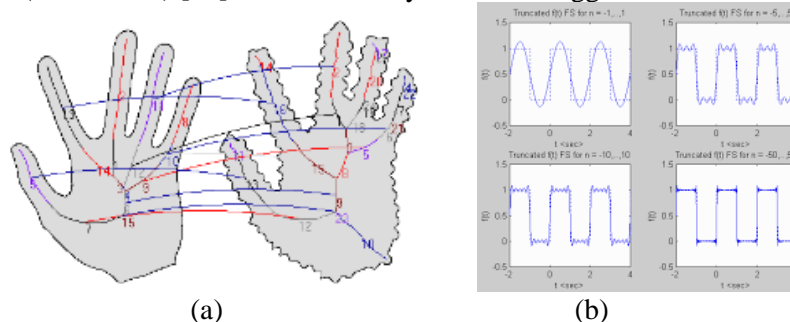
Pada faktor tekstur, Ojala menemukan bahwa dengan menggabungkan histogram tekstur akan relatif lebih sederhana dibandingkan dengan model tekstur tradisional, seperti Gaussian atau Markov [51],[52]. Jafari-Khouzani dan Soltanian-Zadeh mengusulkan fitur tekstur baru berdasarkan *radon* dengan mengubah orientasi pada sisi yang signifikan [53].



Gambar 11. (a) *Gaussian* and (b) *Markov texture*

### 2.4.1 Bentuk

Veltkamp dan Hagedoorn menjelaskan teknik pencocokan bentuk dari perspektif geometri komputasi [54]. Sebe dan Lew mengevaluasi serangkaian bentuk dalam konteks pengambilan gambar [55]. Srivastava menjelaskan beberapa pendekatan baru untuk pembelajaran [56]. Sebastian memperkenalkan gagasan pengenalan bentuk menggunakan grafik *shock* (Gambar 12) [57]. Bartolini menyarankan menggunakan fase *Fourier* [58].



Gambar 12. (a) Bentuk grafik *shock* and (b) Fase *fourier*

### 2.4.2 Audio

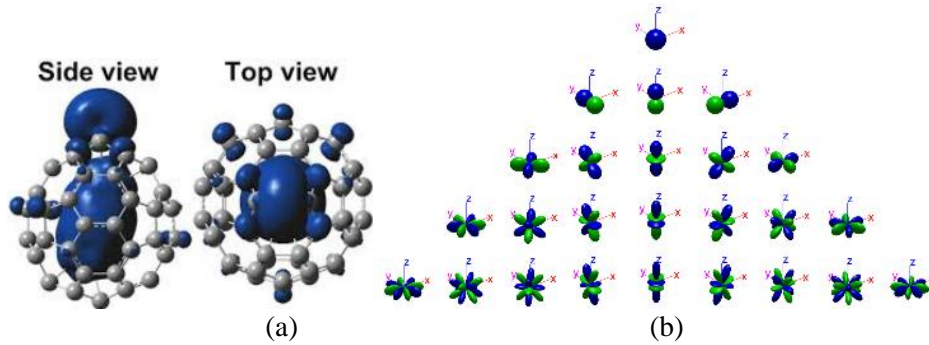
Foote memperkenalkan fitur audio berdasarkan kemiripan lokal [59]. Metode ini melalui perhitungan pada setiap sinyal audio dan metode ini berhasil pada berbagai segmentasi. Bakker dan Lew menyarankan beberapa fitur audio baru melalui perbedaan spektrum frekuensi. Mereka mengevaluasi fitur audio baru dalam konteks pelabelan otomatis sampel sebagai pidato, musik, piano, organ, gitar, mobil, ledakan, atau diam [60].

### 2.4.3 Gambar

Banyak penelitian telah dilakukan pada pencocokan gambar dan strukturnya. Meskipun pendekatan yang digunakan berbeda, sebagian besar metode menggunakan beberapa jenis seleksi titik. Pendekatan melalui deteksi titik dan area ini melalui metode *invariant affine* [61],[62].

## 2.5 Media Baru

Di luar teks, audio, gambar, dan video, ada media baru, seperti model 3D. Assfalg membahas model 3D dengan menggunakan *spin-image* (gambar yang diputar), yang pada dasarnya mengkodekan simpul ke ruang 2D [63]. Funkhouser mengembangkan mesin pencari untuk model 3D berdasarkan pencocokan bentuk menggunakan *spherical harmonics* untuk menghitung langkah-langkah kesamaan diskriminatif yang efektif [64],[65].



Gambar 13. (a) *Spin-Images* dan (b) *Spherical harmonics*

## 2.6 Perangkuman

Ada berbagai macam metode perangkuman informasi multimedia. Spierenburg dan Huijsmans mengusulkan sebuah metode untuk mengonversi *database* gambar ke dalam sebuah film [66]. Metode ini mengelompokkan gambar yang serupa dan menyusunnya secara berurutan, selanjutnya dikonversi ke dalam bentuk video. Hal ini memungkinkan pengguna mendapatkan *gestalt* (persepsi terorganisir) terhadap informasi dari *database* gambar tersebut dalam beberapa menit.

Sundaram mengambil pendekatan yang sama untuk meringkas video. Mereka memperkenalkan gagasan *skim video* yang merupakan video singkat yang terdiri dari adegan informatif dari video asli. Ide dasarnya adalah agar pengguna dapat menerima abstraksi (rangkuman) cerita dalam format video [67]. Snoek mengusulkan beberapa metode untuk meringkas video, seperti pengelompokan berdasarkan kategori dan dalam runtutan waktu [68]. Uchihashi menyarankan metode untuk mengkonversi film ke *strip* kartun dalam gaya Manga dari Jepang [69].



Gambar 14. Konversi ke bentuk manga

## 2.7 Pengindeksan dengan Performa Tinggi

Database multimedia kian bertambah banyak dari ratusan hingga jutaan. Hal ini menyebabkan komputer tidak dapat merespon pencarian dalam jangka waktu cepat. Sistem *database* SQL mulai mengimplementasikan tabel pencarian berperforma tinggi melalui kunci pencarian yang tepat, seperti pada pencarian teks. Suara, gambar, dan video yang disimpan tidak dapat diakses secara efektif. Oleh karena itu, Egas dan Lew mencoba



metode pencarian dengan menggunakan *tree-based indexes* untuk meningkatkan kinerja secara logaritmik [70],[71].

Ye dan Xu menunjukkan bahwa metode vektor kuantisasi dapat digunakan secara efektif untuk pencarian dalam *database* yang besar [72]. Elkwaie dan Kabuka mengusulkan metode berbasis *2-tier signature* untuk pengindeksan *database* gambar yang besar. *Signature 1* menandai sifat dari benda-benda yang ditemukan di gambar, sedangkan *signature 2* menangkap posisi spasial antarobjek. Metode ini telah memberikan peningkatan kerja sebesar 98% [73].

## 2.8 Teknik Evaluasi

Metode evaluasi pada bidang IR dalam satu dekade terakhir melalui TRECVID [74]. Dalam TRECVID, industri dan akademis bersama-sama mengumpulkan, mendiskusikan, dan menyepakati suatu set data multimedia yang spesifik. Banyak tim peneliti IR mencoba untuk membuat berbagai metode untuk IR. Koleksi data multimedia dapat mencakup video dengan naskah, mesin penerjemah non-Inggris, keterangan, metadata, *keyframes*, dan lainnya [75]. *Keyframe* dalam teknik IR merupakan metode yang paling populer dalam sistem video IR [76].

## 3. POTENSI PENELITIAN MIR DI MASA DEPAN

Potensi pada pencarian informasi multimedia cukup luas dan beragam. Di masa yang akan datang, potensi penelitian MIR meliputi peran manusia yang tetap menjadi pusat (aktor) terhadap pencarian informasi, kolaborasi *content* multimedia yang lebih beragam, dan penggunaan kata kunci sederhana (folksonomi).

### 3.1 Metode Berpusat pada Manusia

MIR harus tetap berfokus pada pengguna untuk memberikan pemahaman pola dan keinginan pengguna. Faktor atau fitur interaktif yang baru (seperti penciuman dan pendeteksian ekspresi wajah) saat ini masih diabaikan dan perlu diuji di masa depan untuk memberikan kemungkinan-kemungkinan baru, seperti pendeteksian emosi manusia.

### 3.2 Kolaborasi Multimedia

Menemukan cara yang lebih efektif untuk interaksi antara komputer dan manusia semakin penting karena kondisi dunia ini yang semakin *wireless*. Pengolahan berbagai komponen memberikan peran yang potensial, seperti menyalin dan meringkas hasil rapat dalam menghubungkan suara, nama, dan wajah.

### 3.3 Folksonomi

Otomatisasi pencarian pada *content* multimedia adalah masalah yang sulit. Semua mesin pencari yang ada saat ini menggunakan pendekatan berdasarkan kata kunci sederhana. Pencarian informasi pada *content* multimedia akan lebih berhasil apabila ditambahkan kombinasi antara taksonomi (klasifikasi) dan folksonomi (pemberian *tag* atau kata kunci).

#### 4. KESIMPULAN

Berdasarkan hasil tinjauan literatur, berikut kesimpulan yang menjadi tantangan yang penting bagi komunitas peneliti MIR di masa depan:

1. Pencarian semantik yang menekankan pada pendeteksian objek pada media dengan latar yang kompleks
2. Analisis dan algoritma *retrieval* pada multi-komponen terhadap berbagai media, termasuk teks dan konteks informasi
3. Metode pencarian yang lebih interaktif, penggunaan semantik, dan sistem umpan balik relevansi yang lebih baik
4. Evaluasi algoritma IR melalui penggunaan *data set* berbagai pola.

#### REFERENSI

- [1]. Ballard, D. H., & Brown, C. M., *Computer Vision*, Prentice Hall, New Jersey, USA (1982).
- [2]. Levine, M., *Vision in Man and Machine*, McGraw Hill, Columbus (1985).
- [3]. Haralick, R. M., & Shapiro, L. G., *Computer and Robot Vision*, Addison-Wesley, New York, USA (1993).
- [4]. Flickner, M., et al, *Query by image and video content: the QBIC system*, IEEE Computer, September, 23-32 (1995).
- [5]. Bach, J. R., et al, *Virage image search engine: An open framework for image management*, Proceedings of the SPIE Storage and Retrieval for Still Image and Video Databases, California, USA, 76-87 (1996).
- [6]. Hanjalic, A., et al, *A New Method for Key Frame based Video Content Representation*, Image Databases and Multimedia Search, World Scientific, 97-107 (1997).
- [7]. Smith, J. R., & Chang, S. F., *Visually Searching the Web for Content*, IEEE Multimedia 4(3), 12-20 (1997).
- [8]. Frankel, C., et al, *WebSeer: An Image Search Engine for the World Wide Web*, University of Chicago Technical Report 96-14, University of Chicago, USA (1996).
- [9]. Bliujute, R., et al, *Developing a DataBlade for a New Index*, Proceedings of IEEE International Conference on Data Engineering, IEEE, Sydney, 314-323 (1999).
- [10]. Egas, R., et al, *Adapting k-d Trees to Visual Retrieval*, Proceedings of the International Conference on Visual Information Systems, Amsterdam, 533-540 (1999).
- [11]. Lienhart, R., *Reliable Transition Detection in Videos: A Survey and Practitioner's Guide*, International Journal of Image and Graphics 1(3), 469-486 (2001).
- [12]. HINT.FM, *WebSeer*, FM, <http://hint.fm/seer/> (2009).
- [13]. Fleshmap, *Touch*, Fleshmap, <http://www.fleshmap.com/touch/sorting.html> (2008).
- [14]. Lew, M. S., *Next Generation Web Searches for Visual Content*, IEEE Computer, 46-53 (2000).
- [15]. Lan, T., et al, *From subcategories to visual composites: A multi-level framework for object detection*, IEEE International Conference on Computer Vision (ICCV), pp. 369-376 (2013).
- [16]. Dimitrova, N., *Multimedia Content Analysis: The Next Wave*, Proceedings of the 2nd International Conference on Image and Video Retrieval, London, 9-18 (2003).
- [17]. Lew, M. S., & Denteneer, D., *Fisher Keys for Content Based Retrieval*, Image and Vision Computing 19, 561-566 (2001).
- [18]. Sebe, N., et al, *The State of the Art in Image and Video Retrieval*, Proceedings of the 2nd International Conference on Image and Video Retrieval, London (2003).

- [19]. Jaimer, A., & Sebe, N., *Multimodal Human-computer Interaction: A survey*, Computer Vision and Image Understanding (2006).
- [20]. Rodden, K., et al, *Does organisation by similarity assist image browsing?*, Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, 190-197 (2001).
- [21]. Selmeczi, D., *Manado Tua*. Diakses dari [http://www.divetime.com/photos/Featured/Daniel\\_Selmeczi/Manado\\_Tua\\_6525.html](http://www.divetime.com/photos/Featured/Daniel_Selmeczi/Manado_Tua_6525.html) (2009).
- [22]. Berthouze, N. B. & KATO, T., *Towards a comprehensive integration of subjective parameters in database browsing*, Advanced Database Systems for Integration of Media and User Environments, World Scientific: Singapore, 227-232 (1998).
- [23]. Picard, R. W., *Affective Computing*, MIT Press, Cambridge, USA (2000).
- [24]. Hanjalic, A. & Xu, L-Q., *Affective Video Content Representation and Modeling*, IEEE Transactions on Multimedia, 7(1):171-180 (2005).
- [25]. Sebe, N., et al, *Emotion recognition using a Cauchy naive Bayes classifier*. Proceedings of International Conference on Pattern Recognition, Quebec, 17-20 (2002).
- [26]. Emami, S., *3D Faces From 2D Photos, & Image Processing Difficult Facial Images*, Diakses dari <http://www.shervinemami.info/> (2011).
- [27]. Wang, W., et al, *Image Emotional Classification: Static vs. Dynamic*. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, 6407-6411 (2004).
- [28]. Salway, A., & Graham, M., *Extracting Information about Emotions in Films*, Proceedings of the ACM International Conference on Multimedia, Berkeley, USA, 299-302 (2003).
- [29]. Therrien, C. W., *Decision, Estimation, and Classification*, Wiley, New York, USA (1989).
- [30]. Winston, P., *Artificial Intelligence*, Addison-Wesley, New York, USA (1992).
- [31]. Haralick, R. M., & Shapiro, L. G., *Computer and Robot Vision*, Addison-Wesley, New York, USA (1993).
- [32]. Divvala, S. K., et al, *An empirical study of context in object detection*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1271-1278 (2009).
- [33]. Lew, M. S., & Huijsmans, N., *Information Theory and Face Detection*. Proceedings of the International Conference on Pattern Recognition, Vienna, Austria, 601-605 (1996).
- [34]. Rowley, H., et al, *Human Face Detection in Visual Scenes*, Advances in Neural Information Processing Systems 8 (Proceedings of NIPS), Denver, USA, November, 875-881 (1996).
- [35]. Schneiderman, H. & Kanade, T., *Object Detection Using the Statistics of Parts*, International Journal of Computer Vision 56(3), 151-177 (2004).
- [36]. Chua, T. S., et al, *Detection of human faces in a compressed domain for video stratification*, The Visual Computer 18(2), 121-133 (2002).
- [37]. Vukadinovic, D., & Pantic, M., *Fully automatic facial feature point detection using Gabor feature based boosted classifiers*, IEEE International Conference on Systems, Man and Cybernetics, Vol. 2, pp. 1692-1698 (2005).
- [38]. Lew, M. S., *Next Generation Web Searches for Visual Content*, IEEE Computer, 46-53 (2000).
- [39]. Li, J. & Wang, J. Z., *Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach*, IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1075-1088 (2003).
- [40]. Rautiainen, M., et al, *Detecting Semantic Concepts from Video Using Temporal Gradients and Audio Classification*, Proceedings of the 3rd International Conference on Image and Video Retrieval, London, 260-270 (2003).

- [41]. Rocchio, *Relevance Feedback in Information Retrieval*, The Smart Retrieval System: Experiments in Automatic Document Processing, Ed. Prentice Hall, Englewoods Cliffs (1971).
- [42]. Chen, Y., et al, *One-class SVM for Learning in Image Retrieval*, Proceedings of IEEE International Conference on Image Processing, Thessaloniki, Greece, 815-818 (2001).
- [43]. He, X., et al, *Learning and inferring a semantic space from user's relevance feedback for image retrieval*, Proceedings of the ACM Multimedia. ACM, New York, 343–347 (2002).
- [44]. Yin, P.Y., et al, *Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1536-1551 (2005).
- [45]. BBNC, *Multimedia Analysis Group*, Diakses dari <http://media.au.tsinghua.edu.cn/Research/Multimedia%20Analysis%20Group%20.jsp> (n.d).
- [46]. Lew, M. S., *Principles of Visual Information Retrieval*, Springer, London, UK (2001).
- [47]. Gevers, T., *Color-based Retrieval*, Principles of Visual Information Retrieval, M.S. LEW, Ed. Springer-Verlag, London, 11-49 (2001).
- [48]. Adobe, *The RGB (CMY) Color Model*, Diakses dari [http://dba.med.sc.edu/price/irf/Adobe\\_tg/models/rgbcmy.html](http://dba.med.sc.edu/price/irf/Adobe_tg/models/rgbcmy.html) (2000).
- [49]. Sehbeben, *Color Space*, Diakses dari <http://sehbeben.deviantart.com/journal/Color-Space-409026727> (2013).
- [50]. Sebe, N. & Lew, M. S., *Color Based Retrieval*, Pattern Recognition Letters 22(2), 223-230 (2001).
- [51]. Ojala, T., Pietikainen, M., dan Harwood, D., *Comparative Study of Texture Measures with Classification based on Feature Distributions*, Pattern Recognition 29(1), 51-59 (1996).
- [52]. Haindl, M., & Havlíček, V., *Colour Texture Modelling*, Diakses dari [http://www.ercim.eu/publication/Ercim\\_News/enw44/haindl2.html](http://www.ercim.eu/publication/Ercim_News/enw44/haindl2.html) (2001).
- [53]. Khouzani, K. & Soltanian-Zadeh, H., *Radon Transform Orientation Estimation for Rotation Invariant Texture Analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(6), 1004-1008 (2005).
- [54]. Veltkamp, R.C. & Hagedoorn, M, *State of the Art in Shape Matching. In Principles of Visual Information Retrieval*, Springer-Verlag, London, 87-119 (2001).
- [55]. Sebe, N., & Lew, M.S, *Robust Shape Matching*. In Proceedings of the 1st International Conference on Image and Video Retrieval (2002).
- [56]. Srivastava, A., et al, *Statistical Shape Analysis: Clustering, Learning, and Testing*, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(4), 590-602 (2005).
- [57]. Sebastian, T.B., et al, *Recognition of Shapes by Editing Their Shock Graphs*, IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 550-571 (2004).
- [58]. Bartolini, I., et al, *WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance*, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(1), 142-147 (2005).
- [59]. Foote, J., *Automatic Audio Segmentation Using a Measure of Audio Novelty*. In Proceedings of the IEEE International Conference on Multimedia and Expo, IEEE, Computer Society Press, Los Alamitos, CA, 452–455 (2000).
- [60]. Bakker, E.M. & Lew, M.S, *Semantic Video Retrieval Using Audio Analysis*, In Proceedings of the 1st International Conference on Image and Video Retrieval, London, July 2002, Springer-Verlag, London, 262-270 (2002).
- [61]. Schmid, C., et al, *Evaluation of Interest Point Detectors*, International Journal of Computer Vision, 37(2), 151–172 (2000).

- [62]. Mikolajczyk K. & Schmid, C., *Scale and Affine Invariant Interest Point Detectors*, International Journal of Computer Vision, 60(1), 63–86 (2004).
- [63]. Assfalg, J., et al, *Retrieval of 3D Objects by Visual Similarity*, In Proceedings of the 6th International Workshop on Multimedia Information Retrieval, New York, October 2004, ACM, 77-83 (2004).
- [64]. Funkhouser, T., et al, *A Search Engine for 3D Models*, ACM Transactions on Graphics 22(1), 83-105 (2003).
- [65]. Haas, M., et al, *Relevance Feedback: Perceptual Learning and Retrieval in Biocomputing, Photos, and Video*, In Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, New York, October, 151-156 (2004).
- [66]. Spierenburg, J.A. & Huijsmans, D.P, *VOICI: Video Overview for Image Cluster Indexing*, In Proceedings of the Eight British Machine Vision Conference, Colchester, June (1997).
- [67]. Sundaram, H., et al, *A Utility Framework for the Automatic Generation of Audio-Visual Skims*, Proceedings of the tenth ACM international conference on Multimedia, Juan-les-Pins, France, 189-198 (2002).
- [68]. Snoek, C.G.M., et al, *MediaMill: exploring news video archives based on learned semantics*, In Proceedings of the 13th ACM international conference on Multimedia, Singapore, November, 225-226 (2005).
- [69]. Uchihashi, S., et al, *Video Manga: generating semantically meaningful video summaries*, In Proceedings of the seventh ACM international conference on Multimedia, Orlando, USA, 383-392 (1999).
- [70]. Egas, R., et al, *Adapting k-d Trees to Visual Retrieval*, In Proceedings of the International Conference on Visual Information Systems, Amsterdam, June 1999, 533-540 (1999).
- [71]. Lew, M.S., *Next Generation Web Searches for Visual Content*, IEEE Computer, November, 46-53 (2000).
- [72]. Ye, H. & Xu, G., *Fast Search in Large-Scale Image Database Using Vector Quantization*, In Proceedings of the 2nd International Conference on Image and Video Retrieval, Urbana, July 2003, Springer-Verlag, London, 477-487 (2003).
- [73]. Elkwae, E.A. & Kabuka, M.R., *Efficient Content-Based Indexing of Large Image Databases*, ACM Transactions on Information Systems 18(2), 171-210 (2000).
- [74]. Smeaton, A.F. & Over, P., *Benchmarking the Effectiveness of Information Retrieval Tasks on Digital Video*, In Proceedings of the 2nd International Conference on Image and Video Retrieval, Urbana, July 2003, Springer-Verlag, London, 10-27 (2003).
- [75]. Huijsmans, D.P. & Sebe, N., *How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope*, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2), 245-251 (2005).
- [76]. Pickering, M.J. & Ruger, S., *Evaluation of key-frame based retrieval techniques for video*, Computer Vision and Image Understanding 92(2), 217-235 (2003).