

Pozyskiwanie informacji multimedialnych – bieżące podejścia i zastosowanie

Streszczenie

W pracy przybliżono problematykę związaną z pozyskiwaniem informacji multimedialnych, skupionych wokół danych obrazowych i muzycznych. Zaprezentowane zostały podejścia do pozyskiwania informacji wykorzystujące nie tylko dane opisujące zawartość (metadane) ale przede wszystkim bazujące na charakterystykach tej zawartości. Autor wskazuje główne trendy rozwoju mechanizmów pozyskiwania informacji multimedialnej oraz zastosowanie tych mechanizmów w gotowych rozwiązaniach.

Słowa kluczowe: informacje multimedialne, dane obrazowe, metadane

Wprowadzenie

Dynamiczny rozwój cyfryzacji danych niesie ze sobą praktycznie niekontrolowany wzrost informacji dostępnych w globalnej sieci Internet. Poprzez odpowiednie przetworzenie danych użytkownicy są w stanie pozyskać z nich interesujące informacje. Wyszukiwanie informacji tekstowych, obrazowych czy muzycznych w pewnym ogólnym uproszczonym ujęciu może być podobne. Zarówno zawartość dokumentu tekstowego jak i obrazu czy utworu muzycznego może być opisane za pomocą metadanych (tzn. danych opisujących inne dane, np. tytuł, autor, data powstania dokumentu). Taki zgrubny opis może być niewystarczający (a w niektórych przypadkach nawet nieadekwatny) do opisu zawartości dokumentu, dlatego w nowoczesnych podejściach pozyskiwania informacji multimedialnych bazuje się na wydobyciu cech zawartości wprost z danych dokumentu, które to cechy stanowią charakterystykę opisującą informację.

Trudność w pozyskiwaniu informacji z danych multimedialnych jest związana z formą reprezentacji tych danych. W tabeli 1 zaprezentowano porównanie jawności struktury w ujęciu trzech formatów danych opisujących informację – tekstowych, muzycznych oraz obrazowych.

Pozyskiwanie informacji można obecnie uogólnić do trzech głównych metodyk. Pierwsza metodyka polega na samodzielnym przeszukiwaniu repozytoriów danych cyfrowych przez użytkownika.

Druga metodyka obecnie bardzo szeroko stosowana, polega na pozyskiwaniu informacji na podstawie opisu tekstowego, skrótowo przedstawiającego zawartość dokumentu, obrazu, nagrania audio lub wideo. W tej grupie metod informacja tekstowa (pełny tekst, dokumenty tekstowe czy metadane dokumentów multimedialnych) jest indeksowana przez repozytoria, tak, aby użytkownik mógł sprawnie odszukać interesujące go treści, podając słowa kluczowe lub pojęcia.

Tabela 1. Porównanie jawności struktury danych tekstowych, muzycznych i obrazowych.

Informacja tekstowa	Informacja muzyczna	Informacja obrazowa	Jawność struktury
Tekst ze znacznikami (HTML)	Notacja muzyczna (nuty)	Złożone obiekty, sceny z opisem ich rozmieszczenia i budowy (np. pliki scen i modeli 3D)	maksymalna
Niesformatowany tekst (TXT)	Zdarzenia czasowe (MIDI)	Obraz z częściowym opisem cech (np. wykorzystujący MPEG-7)	średnia
Audio (mowa)	Audio (CD, pliki MP3)	Obraz (plik JPG)	minimalna

Źródło: opracowanie własne na podstawie ¹

Ostania metodyka polega na pozyskiwaniu informacji bazując na zawartości. Takie podejście pozwala użytkownikom wyszukiwać informacje multimedialne w zakresie odpowiadającym faktycznej zawartości obrazu, audio i wideo². Oczywiście może się wydawać, że metodyka wyszukiwania informacji w oparciu o opis tekstowy jest wystarczająca, to jednak istnieją pewne odstępstwa od tej reguły.

Pozyskiwanie informacji bazujące na zawartości jest niezbędne tam, gdzie dokonanie opisu tekstowego jest utrudnione lub niemożliwe. Przykładem niech będzie zdjęcie pewnego rzadkiego gatunku owada, o którym należy znaleźć informacje. W podejściu opartym na opisie tekstowym należało by zasięgnąć porady entomologa, który powinien scharakteryzować owada poprzez opis tekstowy i wykorzystać odpowiednie słowa kluczowe do wyszukania publikacji traktujących o tym owadzie. W podejściu wykorzystującym zawartość (w tym przypadku obrazu) należałoby przesłać zdjęcie owada do repozytorium, które przy wykorzystaniu metod pozyskiwania informacji bazujących na zawartości, zwróci listę wyników w postaci obrazów podobnych do podanego ze szczegółowymi opisami owadów znajdujących się na obrazach.

Podobnie przy wykorzystaniu tej metodyki można wyszukać kolekcję muzyczną czy pojedynczy utwór, posiadając jedynie fragment szukanej kompozycji w postaci pliku audio. Co więcej, przy braku informacji o autorze czy tytule kompozycji, wyszukiwanie w oparciu o zawartość jest w tym przypadku dużo prostsze, niż przy wykorzystaniu informacji tekstowych opisujących poszukiwany fragment utworu.

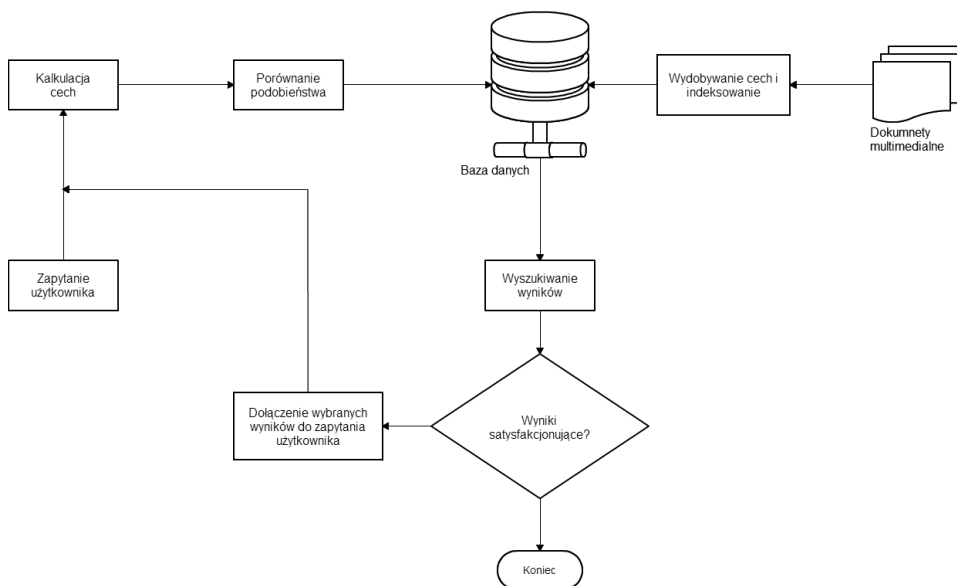
W obecnej chwili sieć WWW dostarcza nowych wyzwań dla tradycyjnego tekstowego wyszukiwania informacji. Chociaż dzisiejsze repozytoria cyfrowe są dostępne na całym świecie, to jednak użytkownicy różnych języków i kultur mogą nie być w stanie skutecznie przeszukiwać takie zbiory za pomocą słów kluczowych. Wyszukiwanie informacji bazujące na zawartości znacznie zwiększają dostępność repozytoriów cyfrowych i to prawdopodobnie jest główny powód gwałtownego rozwoju tego obszaru badań w ciągu ostatniej dekady. W idealnym ujęciu, system wyszukiwania informacji bazującego na zawartości może „znać” powiązania semantyczne multimedialnych danych, takich jak obiekty i ich kategorie. Jednakże trudnością w obecnej technologii komputerowej jest wyodrębnianie z danych multimedialnych wysoko poziomowych lub semantycznych właściwości informacji. Większość projektów nadal koncentruje się na wykorzystaniu cech niższego szczebla, takich jak kolor, faktura czy kształt.

¹ D. Byrd, T. Crawford, Problems of music information retrieval in the real world, *Information Processing and Management*, Vol. 38, Nr 2, 2002, s. 249–272.

² O. Marques, Furht B, *Content-based Image and Video Retrieval*, 2002, Norwell, Mass: Kluwer.

W ogólności system pozyskiwania informacji opartego na zawartości można opisać za pomocą kilku zdań. Po pierwsze, w odniesieniu do każdego pliku multimedialnego w bazie danych, wydobywana jest pewna cecha informacji (np. kolor, ruch czy wysokość dźwięku), która jest następnie indeksowana i zapisywana. Po drugie, gdy użytkownik tworzy zapytanie, cechy informacji tego zapytania są przekształcane (obliczane) do postaci wektorowej. Wreszcie, system porównuje podobieństwo pomiędzy wektorem cech tego zapytania i danymi multimedialnymi (ich cechami) w bazie danych, a następnie zwraca najlepiej pasujące rekordy (wyniki np. w postaci listy). Jeżeli użytkownik nie jest zadowolony z rezultatów, może uściślić wyszukiwanie poprzez wybór najbardziej odpowiednich wyników i dołączenie ich do zapytania, a następnie powtórzyć wyszukiwanie nowej informacji. Schemat tego działania przedstawiono na rysunku 1.

Rysunek 1. Ogólny proces pozyskiwania informacji opartego na zawartości



Źródło: opracowanie własne

W dalszych sekcjach rozpatrzono niektóre istniejące techniki pozyskiwania (wydobywania) informacji bazujące na zawartości, związane z najpopularniejszymi formatami multimediiów (obraz, audio i video). Wskazano również ich ograniczenia i trendy rozwoju.

Pozyskiwanie informacji obrazowych bazujące na zawartości – systemy CBIR

Przez ostatnich kilka lat pojawiło się wiele systemów CBIR (ang. Content-Based Image Retrieval), które albo bazowały na wcześniejszych rozwiązaniach, albo wskazały nowe kierunki rozwoju. Systemy te posiadają pewną wspólną cechę, którą jest pozyskanie

globalnych cech obrazu, takich jak kolor, kształt i tekstura w pierwszej fazie procesu wydobywania informacji³.

Ważnym elementem systemu CBIR jest moduł ekstrakcji cech z obrazu. Tematyka ta jest na tyle rozległa, że w pracy przedstawiono jedynie kilka podejść charakterystycznych dla tego obszaru badań.

W pracy Denga⁴ i innych zaproponowano rozwiązanie bazujące na deskrytorze dominującego koloru obszaru obrazu. Deskrytor jest indeksowany w przestrzeni trójwymiarowej wraz z procentowym pokryciem koloru obszarów. Podejście wydaje się być bardziej wydajne obliczeniowo w porównaniu do tradycyjnych histogramów koloru.

Hadjidemetriou⁵ i inni zaproponowali wykorzystanie wielorozdzielczego histogramu przechwyconej informacji obrazu przestrzennego w celu efektywnego pozyskiwania obrazów teksturowanych, zachowując przy tym typowe zalety rozwiązań bazujących na histogramach.

Zestaw deskrytorów koloru i tekstury, rygorystycznie testowany przed włączeniem go w standard MPEG-7 i dobrze nadający się do wykorzystania z obrazami naturalnymi i materiałem video, został przedstawiony w pracy Manjunatha⁶ i innych. Obejmuje on deskryptory bazujące na histogramie, deskryptory dominującego koloru (DCD), przestrzennych deskrytorów kolorów oraz deskrytorów tekstury.

Przy opisie technik wydobywania cech z obrazów, bazujących na deskrytorach cech, wspomnieć należy o pracy Forczmańskiego⁷, w której zaprezentowana została aplikacja oferująca możliwość wyszukiwania obrazów podobnych do wejściowego na podstawie cech wybranych przez użytkownika. Użytkownik na wejściu programu podaje przykładowy obraz, będący wzorcem wyszukiwania. Wynikiem jest zbiór obrazów posortowanych w kolejności podobieństwa (od najbardziej podobnego). W aplikacji zaimplementowano kaskadowy sposób łączenia deskrytorów, który pozwala na intuicyjne tworzenie różnorodnych schematów wyszukiwania.

Kształt jest kluczowym atrybutem w wydzielonych segmentach obrazu, oraz skuteczną i niezawodną reprezentacją, która odgrywa ważną rolę w pozyskiwaniu informacji. Analizę porównawczą najczęściej wykorzystywanych deskrytorów kształtu, w generalnym problemie analizy kształtu (z ang. General Shape Analysis Problem) przedstawił Frejlichowski⁸. Do przytoczonej analizy rozwiązań, należy dodać komentarz, w postaci złożoności obliczeniowej deskrytorów Fouriera oraz inwariantów momentów, zaprezentowanych

³ R. Datta, J. Li, J.Z. Wang, Content-based image retrieval: approaches and trends of the new age, In Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in Conjunction with ACM International Conference on Multimedia, H. Zhang, J. Smith, and Q. Tian, 2005, eds. New York: ACM.

⁴ Y. Deng Y., B.S. Manjunath, C. Kenney, M. S. Moore, i H. Shin, An Efficient Color Representation for Image Retrieval, IEEE Trans. Image Processing, Vol.10, Nr 1, 2001, s. 140–147.

⁵ E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, Multiresolution Histograms and Their Use for Recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 26, Nr 7, 2004, s. 831–847.

⁶ B.S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," IEEE Trans. Circuits and Systems for Video Technology, Vol. 11, Nr 6, 2001, s. 703–715.

⁷ P. Forczmański, P. Szeptycki, Indeksowanie dużych zbiorów obrazów, Roczniki Informatyki Stosowanej Wydziału Informatyki PS, Nr 8, Metody informatyki stosowanej w technice i technologii, Szczecin 2005.

⁸ D. Frejlichowski, An Experimental Comparison of Seven Shape Descriptors in the General Shape Analysis Problem. In: A. Campilho and M. Kamel (Eds.): ICIAR 2010, Part I, Lecture Notes in Computer Science, vol. 6111, 2010, s. 294–305.

w tej pracy. Oczywiście, przy wykorzystaniu pewnych obliczeń wstępnych (prekalkulacji), można uzyskać rozwiązania z obszaru systemów czasu rzeczywistego.

Rysunek 2. Główne okno programu Fast Image Finder



Źródło: P. Forczmański, P. Szeptycki, Indeksowanie dużych zbiorów obrazów, Roczniki Informatyki Stosowanej Wydziału Informatyki PS, Nr 8, Metody informatyki stosowanej w technice i technologii, Szczecin 2005

Pomimo iż rozwiązania wykorzystywane w systemach CBIR, bazujące na właściwościach obrazu, są badane przez wiele lat, to nowym powstającym kierunkiem badań w tych systemach jest automatyczne rozpoznawanie pojęcia (z wizualnych cech obrazu) i automatyczne adnotacje. Jest to wyzwanie przede wszystkim ze względu na występującą lukę semantyczną, pomiędzy niskopoziomowymi cechami wizualnymi a wysokim poziomem pojęć. A więc głównym celem praktycznym systemów CBIR (w tym ujęciu) jest odkrycie obrazów odnoszących się do danego pojęcia, przy braku wiarygodnych i rzetelnych metadanych.

Adnotacje pozwalają na wyszukiwanie obrazów przy użyciu tekstu. W tym przypadku automatyczne adnotacje mogą być bardziej praktyczne dla dużych zbiorów danych. Jeżeli wynik automatycznego mapowania pomiędzy obrazami i słowami można uznać za wiarygodny, to tekstowe wyszukiwanie obrazów może być semantycznie bardziej znaczące niż CBIR. Próba „zrozumienia” obrazu odbywa się w fazie automatycznego rozpoznawania konceptu (pojęcia). Z tego względu adnotacje można traktować jako podzbiór wykrywania pojęcia, tzn. obrazy odnoszące się do tego pojęcia, mogą być językowo różnie opisane na podstawie konkretnego przykładu pojęcia. Rodzi się więc pytanie, czy cechy wizualne obrazów mogą przekazać jakiegokolwiek informacje na temat własnego pojęcia?

Vailaya⁹ i inni zaprezentowali detekcję (z dużą dokładnością) pojęcia poprzez nadzorowaną klasyfikację prostych pojęć, takich jak miasto, krajobraz, zachód słońca i las. W pracy

⁹ A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, “Image Classification for Content-Based Indexing,” IEEE Trans. Image Processing, Vol 10, Nr (1), 2001, s. 117–130.

Donga i Bhanu¹⁰ zaprezentowano podejście uczenia pojęć z wykorzystaniem sprzężenia zwrotnego z użytkownikiem, dynamicznie zmieniającej się bazy danych oraz gaussowskich modeli mieszanych (z ang. Gaussian Mixture Models, GMM). Do automatycznych adnotacji obrazów ze zbiorem kilkuset słów, wykorzystywano również dwuwymiarowe wielorozdzielcze ukryte modele Markova (z ang. Hidden Markov Models, HMM)¹¹.

Chociaż CBIR jest głównym obszarem badawczym i z ma najdłuższą historię w dziedzinie pozyskiwania informacji bazującego na zawartości, to jednak istnieje wiele modeli, produktów oraz bieżących projektów uzupełniających podane tu przykłady. Metody CBIR wykorzystywane są w szerokim aspekcie możliwości: od wyszukiwania obrazów w repozytoriach cyfrowych (np. bibliotekach cyfrowych), poprzez diagnozę medyczną opartą na analizie obrazu medycznego po zadania identyfikacji obiektów na mapach lub inne rozwiązania stosowane np. przez służby wojskowe i policję.

Pozyskiwanie informacji muzycznej bazujące na zawartości – systemy CBMIR

W porównaniu do systemów CBIR, pozyskiwanie informacji audio bazujące na zawartości (z ang. Content-Based Audio Retrieval, CBAR) jest stosunkowo młodą gałęzią rozwoju. W ogólności, istniejące rozwiązania CBAR rozpoczynają się od analizy klipu audio. Przykładem takiej analizy jest wyodrębnianie podstawowych elementów audio, takich jak czas trwania, wysokość dźwięku, amplituda czy przepustowość. W systemach CBAR analizowane są zarówno materiały zawierające odgłosy zwierząt, ludzi, pracy elementów mechanicznych i elektronicznych (wykrywanie usterek tych elementów) jak również materiały audio zawierające nagrania muzyczne.

Z punktu widzenia popularności, dominującą grupą w tej dziedzinie są systemy pozyskiwania informacji muzycznej bazujące na zawartości (z ang. Content-Based Music Information Retrieval, CBMIR) i to właśnie ta grupa rozwiązań będzie przedstawiona w tej sekcji.

Dużą grupę użytkowników systemów MIR stanowią użytkownicy mający określone oczekiwania muzyczne w sferze przeszukiwania kolekcji muzycznych. Jednak wraz ze wzrostem aktywności konsumentów w dziedzinie muzyki cyfrowej, pojawiają się nowe możliwości badań, przy użyciu dużych kolekcji muzyki, trendów i wzorców pojawiających się w utworach muzycznych. Systemy rozpoznawania trendów w sprzedaży online muzyki istnieją już w postaci komercyjnej¹², jak również systemy, które wspomagają badania muzykologiczne nad ewolucją muzyki przez korpus nut zachodniej muzyki klasycznej i dostępnych nagrań muzyki klasycznej¹³.

Istnieją trzy główne grupy odbiorców, których można nazwać beneficjentami systemów MIR: organy (podmioty) branżowe zaangażowane w nagrywanie, agregację i rozpowszechnianie muzyki; użytkownicy końcowi, chcą znaleźć muzykę i używać jej indywidualnie; profesjonalści: wykonawcy muzyki, nauczyciele, muzykolodzy, prawnicy specjalizujący się w prawach autorskich i producenci muzyki.

¹⁰ A. Dong and B. Bhanu, "Active Concept Learning for Image Retrieval in Dynamic Databases," Proc. IEEE International Conference on Computer Vision, 2003.

¹¹ J. Li, J. Z. Wang, "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 25, Nr (9), 2003, s. 1075–1088.

¹² A. Jha, "Music machine to predict tomorrow's hits", Guardian, Styczeń 17, 2006. [dostęp on-line: 22.02.2013] <http://www.guardian.co.uk/technology/2006/jan/17/news.science>.

¹³ E. Clarke, N. Cook, *Empirical Musicology: Aims, Methods, Prospects*. New York: Oxford Univ. Press, 2004.

Obecnie najpopularniejszą metodą dostępu do muzyki jest jej wyszukiwanie za pomocą tekstowych metadanych. Metadane mogą nieść ze sobą wyraziste i bogate znaczenie, więc istnieje wiele scenariuszy rozwiązań, w których takie podejście jest wystarczające. Obecnie większość usług pobierania muzyki używa tylko podejścia bazującego na metadanych i osiągnęło poziom sukcesu komercyjnego. Jednak w przypadku, kiedy katalogi muzyczne stają się bardzo duże (większe niż setki tysięcy ścieżek muzycznych), niezmiernie trudno zachować się spójność ekspresyjnych deskryptorów metadanych, tworzonych przez wiele osób z różnym podejściem do pojęć wyrażanych przez te metadane.

Serwisy internetowe social mediów otwierają zadanie opisu zawartości muzycznej szerokim grupom społecznościowym użytkowników, którzy później wymieniają się tymi informacjami między sobą. Jest to znak rozpoznawczy technologii i ideologii Web 2.0. Miliony użytkowników portali takich jak MySpace, Flickr i YouTube wykazuje zachowania grupowe, które naturalnie przyciąga ich do pojęcia części portalu jaką jest kategoria lub grupa, w obrębie której wymieniają się zainteresowaniami, dzięki czemu z większą dozą prawdopodobieństwa mogą znaleźć interesujące treści, umieszczone przez użytkowników o podobnych gustach.

W uzupełnieniu do systemów bazujących na metadanych, istnieją rozwiązania bazujące na zawartości, które pomagają użytkownikom odnaleźć interesujące ich treści, nawet wtedy, gdy nie wiedzą konkretnie czego mają szukać. Przykładem takiego rozwiązania jest system Shazam¹⁴ (shazam.com), który potrafi identyfikować nagrania z próbki pobranej z telefonu komórkowego, nagranej w klubie muzycznym lub zatłoczonym barze, a następnie dostarczyć informacji o artyście, albumie czy tytule piosenki wraz z danymi o lokalizacjach najbliższych punktów zakupu tego nagrania lub bezpośrednim odnośnikiem do opcji zakupu/pobrania tego utworu.

W przemyśle nagraniowym, firmy używają systemów wykorzystujących symboliczne informacje zawartości muzycznej (melodia, akordy, rytm czy słowa), aby zanalizować potencjalny wpływ pracy na rynek. Serwisy takie jak Hit Song Science (fundacji Polyphonic HMI) i Platinum Blue Music Intelligence używają takich informacji z wykorzystaniem metod sztucznej inteligencji, aby tworzyć rekomendacje o nowych wydaniach (utworach, potencjalnych hitach).

Metody wykorzystywane w systemach CBMIR są również wykorzystywane w zupełnie odmiennej i dość niszowej gałęzi nauki jaką są systemy algorytmicznego komponowania muzyki (tworzenie muzyki przez maszynę). W tym aspekcie metody MIR są wykorzystywane do pozyskiwania i klasyfikacji charakterystycznych motywów muzycznych (fragmentów utworu), właściwości symbolicznych utworów (rytm, harmonia, melodia), używanych do komponowania nowego utworu muzycznego. Takie systemy ogólnie można podzielić na: systemy komponowania wykorzystujące jawne reguły kompozycji (zdefiniowane wcześniej przez twórcę) oraz systemy komponowania wykorzystujące niejawne reguły kompozycji (pozyskane z przykładowych utworów muzycznych^{15,16} jako wzorce dla nowych kompozycji).

¹⁴ A. Wang, "The Shazam music recognition service", *Com. ACM*, Vol. 49, Nr. 8, 2006, s. 44–48.

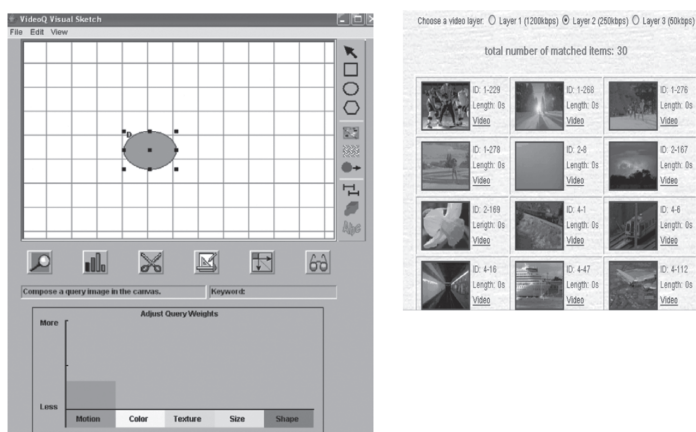
¹⁵ Ł. Mazurowski, "An algorithmic composition – the chosen aspects", *Przegląd Elektrotechniczny (Electrical Review)*, R.88 10b, 2012, s. 243–246,

¹⁶ Ł. Mazurowski, "Computer models for algorithmic music composition", In *Proceedings of the FedCSIS 2012 (MMAF)*, IEEE Xplore Digital Library, s. 733–737.

Pozyskiwanie informacji wideo bazujące na zawartości – systemy CBVR

Systemy pozyskiwania informacji wideo bazującego na zawartości (z ang. Content-Based Video Retrieval, CBVR) jest nowszym zagadnieniem niż systemy CBIR czy CBMIR, częściowo dlatego, że technologia digitalizacji dla wideo pojawiła się później niż dla obrazu i dźwięku. Po tym jak serwisy typu YouTube i Google Video stawały się co raz bardziej popularne, znajdowanie żadanego klipu wideo w sposób skuteczny stało się problemem. Wyszukiwanie wykorzystujące niektóre cechy wideo, takie jak ruch obiektu czy tekstura, może być dobrym uzupełnieniem tradycyjnego tekstowego wyszukiwania zawartości wideo.

Rysunek 3. Komponowanie zapytania użytkownika (po lewej) oraz wyniki wyszukiwania przykładowego zapytania (po prawej) – system VideoQ



Źródło: G. Wan, Z. Liu, Content-Based Information Retrieval and Digital Libraries, Information Technology & Libraries, Vol. 27, Nr 1, 2008., s. 41–47.

Jednym z pierwszych rozwiązań dotyczących systemów CBVR był system VideoQ stworzony przez Changa¹⁷ i innych, który pozwalał użytkownikowi na wyszukiwanie wideo w oparciu o bogaty zestaw cech wizualnych oraz relacji czasoprzestrzennych. Plik wideo były przechowywane w bazie danych w postaci plików MPEG. Poprzez interfejs WWW, użytkownik mógł formułować zapytanie-scenę jako kolekcję obiektów z różnymi atrybutami, włączając ruch, kształt, kolor i teksturę. Po sformułowaniu zapytania, jest ono wysyłane na serwer, które zawiera kilka różnych baz danych dla różnych cech zawartości. Po stronie serwera obliczane są podobieństwa pomiędzy cechami każdego określonego w zapytaniu obiektu a cechami obiektów z baz danych. Zwracana jest lista klipów wideo, bazująca na wartościach podobieństwa. Dla każdego klipu wideo są następnie dynamicznie

¹⁷ S. Chang, et al., "VideoQ: an automated content based video search system using visual cues". In Proceedings of the 5th ACM International Conference on Multimedia, E. P. Glinert, et al., eds. New York: ACM, 1997.

wyodrębniane klatki kluczowe z wideo baz danych, które zwracane są do przeglądarki. Dopasowane obiekty są podświetlone w zwracanej klatce kluczowej. Użytkownik może interaktywnie obejrzeć te pasujące klipy wideo, klikając w klatkę kluczową.

Systemy CBVR są również wykorzystywane w zagadnieniach śledzenia trajektorii ruchu obiektu i jego identyfikacji na scenie¹⁸. Z tym obszarem badań wiąże się wiele problemów związanych z eliminowaniem wpływu oświetlenia sceny i złożoności jej tła na identyfikację obiektu i predykcję jego ruchu, a także eliminacja zaszumienia sceny, wprowadzającego błędy identyfikacji i predykcji.

Rysunek 4. Serie klatek filmowych w warunkach zachmurzonego otoczenia i paskowanym tłem sceny. Klatki bieżące oraz klatki wyodrębnione (czarne tło) ze zidentyfikowanym obiektem.



Źródło: C. Y. Wen, L. F. Chang, H. H. Li., „Content based video retrieval with motion vectors and the RGB color model,” *Forensic Science Journal*, Vol. 6, Nr.2, s. 1–36, 2007.

Na rysunku 4 przedstawiono przykład identyfikacji obiektu (człowieka) w warunkach zachmurzenia na paskowanym tle sceny (na parkingu samochodowym). W prezentowanym systemie wykorzystano wektory ruchu oraz model przestrzeni barw RGB do pozyskiwania interesujących ruchomych obiektów z nagrań video na potrzeby policji. W opracowanej metodzie wprowadzono etapy eliminacji tła oraz cieni tworzonych przez obiekty sceny, a także eliminację szumu oraz ekstrakcję kształtu i konturu obiektów.

¹⁸ C. Y. Wen, L. F. Chang, H. H. Li., „Content based video retrieval with motion vectors and the RGB color model,” *Forensic Science Journal*, Vol. 6, Nr.2, s. 1–36, 2007.

Podsumowanie

Nie ma wątpliwości, że pozyskiwanie informacji bazujące na zawartości jest nowym trendem w rozwoju wielu usług oferowanych przez współczesne media i jest ważnym elementem uzupełniającym tradycyjne tekstowe technologie pozyskiwania informacji. Idealny system pozyskiwania informacji bazującego na zawartości, może semantycznie rozumieć informacje zawarte w repozytorium cyfrowym, i dostarczać użytkownikom pożądaných danych na podstawie analizy składki zapytania. Jednakże rozumienie maszynowe semantycznej informacji nadal pozostaje trudne do zrealizowania przez lukę semantyczną istniejącą między niskopoziomowymi cechami zawartości a ich znaczeniami (pojęciami) semantycznymi. Najnowsze projekty badawcze w tej dziedzinie, skupiają się na rozumieniu i pozyskiwaniu tych niskopoziomowych cech lub właściwości fizycznych treści multimedialnych. Oczywiście, jeśli im bardziej będą się rozwijać pokrewne dyscypliny takie jak widzenie maszynowe czy sztuczna inteligencja, tym więcej badań będzie robionych w ramach pozyskiwania bazującego na właściwościach wysokopoziomowych.

Ponadto większość projektów z tej dziedziny ma tendencję do rozpraszania usług w ramach sieci globalnej Internet (Flickr, YouTube czy Google Video), szczególnie w dobie Web 2.0. Naturą tych systemów jest zapewnienie jak najlepszej możliwości wyszukiwania informacji przez użytkowników końcowych, dlatego bardzo ważnym jest skupienie się na rzeczywistych potrzebach użytkowników oraz możliwości wykorzystania przez nich nowych narzędzi wyszukiwania. Zaskakującym jest fakt, że wykonanych zostało tak mało testów użyteczności dla większości projektów CBIR, CBAR, CBMIR czy CBVR. Takie badania powinny być przeprowadzone, zanim dane rozwiązanie zostanie powszechnie przyjęte.

Abstract

The problems connected with the retrieval of multimedia information concentrated on image and music data are described in the work. The approaches to retrieving information not only by means of the data that describes the content (metadane) but also based on the characteristics of this content are presented. The author points out the major trends in the development of the multimedia information retrieval mechanisms and the usage of these mechanisms in the already existing solutions.