

06171 Abstracts Collection

Content-Based Retrieval

— Dagstuhl Seminar —

Tim Crawford¹ and Remco C. Veltkamp²

¹ Goldsmiths College - London, GB

`t.crawford@gold.ac.uk`

² Utrecht Univ., NL

`remco.veltkamp@cs.uu.nl`

Abstract. From 23.04.06 to 28.04.06, the Dagstuhl Seminar 06171 “Content-Based Retrieval” was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Content-based retrieval, indexing, video analysis, music similarity, shape similarity

06171 Executive Summary – Content-Based Retrieval

The emphasis of the fourth seminar in the Dagstuhl Seminar series on "Content-Based Retrieval" lies on the PERCEPTUAL and COGNITIVE aspects of all kinds of content-based retrieval. Fundamental questions such as the role of perception and cognition in feature extraction, pattern similarity rating, indexing large collections etc. play an important role.

Keywords: Content-based retrieval, indexing, video analysis, music similarity, shape similarity

Joint work of: Crawford, Tim; Veltkamp, Remco C.

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/757>

Efficient multi-step query processing for EMD-based similarity

Ira Assent (RWTH Aachen, D)

Similarity search in large multimedia databases requires efficient query processing based on suitable similarity models. Similarity models consist of a feature extraction step as well as a distance defined for these features, and they demand an efficient algorithm for retrieving similar objects under this model. In this work, we focus on the Earth Movers Distance (EMD), a recently introduced similarity model which has been successfully employed in numerous applications and has been reported as well reflecting human perceptual similarity. As its computation is complex, the direct application of the EMD to large, high-dimensional databases is not feasible. To remedy this and allow users to benefit from the high quality of the model even in larger settings, we developed various lower bounds for the EMD to be used in index-supported multistep query processing algorithms. We prove that our algorithms are complete, thus producing no false drops. We also show that it is highly efficient as experiments on large image databases with high-dimensional features demonstrate.

Keywords: Content-based retrieval, indexing, multimedia databases, efficiency, similarity

Joint work of: Assent, Ira; Seidl, Thomas

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/649>

From pixels to semantics: methods for analysing image content

Ronen Basri (Weizmann Inst. - Rehovot, IL)

This talk will review several methods for analyzing image content, including (1) hierarchical image segmentation, (2) accounting for variability in lighting and pose, and (3) example-based 3D shape reconstruction.

Muugle: A Modular Music Information Retrieval Framework

Martijn Bosma (Utrecht University, NL)

Muugle (Musical Utrecht University Global Lookup Engine) is a user-centered modular framework that allows the comparison of different MIR techniques and usability studies. A system overview and a discussion of a pilot usability experiment are given. A demo version of the framework can be found on <http://give-lab.cs.uu.nl/muugle>.

Keywords: Music Information Retrieval Framework

Efficient Approximate Audio Matching by Hashing

Michael Casey (Goldsmiths College - London, GB)

In this paper we present methods for identifying derivative works in large audio collections, that is, works that contain audio passages that resemble passages from a source work, or set of source works. In our application, resemblance is approximate, we do not look for exact matches of the signal.

This is because derivative works do not simply contain “samples” of the signal of an original work, but instead use one aspect of the source, such as a vocal passage, and remix it with new percussion and instrumental audio. Only a very small part of the source work might be used for the derivative work, so any method used to identify derivative must identify sources in a completely new context, this is called partial containment. Hence identification of derivative works is partial containment of approximately matching audio. Our solution uses audio shingling, a novel method based on techniques for near-duplicate elimination in document search. We describe the steps for constructing and comparing audio shingles, and for identifying partial containment using scalable hashing algorithm for nearest neighbour retrieval.

Keywords: Music similarity, audio shingling, nearest neighbours in high dimensions

After the search is over ... the work begins

Tim Crawford (Goldsmiths College - London, GB)

A narrative and discursive discussion of some adventures in musicology using a simple MIR system.

Keywords: Music information retrieval, musicology, OMRAS, harmonic modeling

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2006/747>

Musicream: Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces

Masataka Goto (AIST - Ibaraki, J)

This paper describes a novel music playback interface, called Musicream, which lets a user unexpectedly come across various musical pieces similar to those liked by the user.

With most previous “query-by-example” interfaces used for similarity-based searching, for the same query and music collection a user will always receive the same list of musical pieces ranked by their similarity and opportunities to encounter unfamiliar musical pieces in the collection are limited. Musicream facilitates active, flexible, and unexpected encounters with musical pieces by providing four functions: the music-disc streaming function which creates a flow of many musical-piece entities (discs) from a (huge) music collection, the similarity-based sticking function which allows a user to easily pick out and listen to similar pieces from the flow, the meta-playlist function which can generate a playlist of playlists (ordered lists of pieces) while editing them with a high degree of freedom, and the time-machine function which automatically records all Musicream activities and allows a user to visit and retrieve a past state as if using a time machine. In our experiments, these functions were used seamlessly to achieve active and creative querying and browsing of music collections, confirming the effectiveness of Musicream.

Keywords: Music interface, Music player, Music-collection browser, Query-by-example, Playlist

Full Paper:

<http://staff.aist.go.jp/m.goto/Musicream/>

See also: Masataka Goto and Takayuki Goto: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), pp.404-411, September 2005.

Paradigm Shifts in Video Content Analysis Needed: The Why’s and How’s of Generic VCA Solutions

Alan Hanjalic (Delft Univ. of Technology, NL)

In the first part of this talk the past and current developments in the field of video content analysis (VCA) are discussed. A number of statements are made emphasizing the need for fundamental VCA concepts and principles, based on which the VCA solutions can be developed that are characterized by more robustness, reliability and much wider applicability than the state-of-the-art VCA solutions. In particular, the “curse of domain knowledge” and the lack of creative, non-conventional, out-of-the-box ideas addressing the true nature of the VCA problems to be solved, are identified as the main bottlenecks preventing the further growth in the VCA research field.

In the second part of the talk the research at Delft University of Technology is presented, that aims at a generic framework for affective video content characterization. The framework is based on the valence-arousal paradigm that is used to uncouple the feature measurements from the semantics (affect) inference and

provides the possibility to elegantly bridge the semantic gap without the need for explicitly modeling the abstract and ambiguous affect categories as required by the classical approaches based on supervised classification. Application potential of the proposed framework is illustrated in the context of personalized video content delivery and video highlights extraction.

Keywords: Video content analysis, affective computing

See also: A. Hanjalic: Extracting Moods from Pictures and Sounds: Towards Truly Personalized TV, IEEE Signal Processing Magazine, March 2006

Some Principles for Constructing Robust and Semantically Meaningful Audio Features

Frank Kurth (Universität Bonn, D)

Motivated by two applications from the field of music information retrieval (MIR), audio identification and audio matching, we address some general principles for constructing robust and, at the same time, semantically meaningful audio features.

Whereas audio identification aims at identifying a short excerpt (let's say of about 10-30 seconds of duration) of audio as being part of a particular audio recording taken from a particular CD, audio matching aims at automatically retrieving all musically similar excerpts in all interpretations of the underlying pieces of music, which are contained in the database. Audio matching may hence be considered as a semantically advanced retrieval problem.

In our talk, we first consider an approach to feature design based on calculating time-domain statistics for the well-known class of pitch-based chroma features. This allows us to successfully approach the problem of audio matching for the case of western classical music. Subsequently, we adapt the chroma-approach to extract robust tempo-related features resulting in the concept of a cyclic beat spectrum (CBS). Using a suitable feature set derived from the CBS, we propose an approach to robust time-scale invariant audio identification.

Keywords: Chroma, CENS, beat spectrogram, cyclic beat spectrum, CBS, audio matching, time-scale invariant audio identification

Partial Shape Similarity Using Edge Groups Detected by Extended EM

Rolf Lakaemper (Temple University - Philadelphia, USA)

The talk's first part will present an extension to global similarity measures on polygonal boundaries to discover partial matches.

The presented partial similarity measure emphasizes the common parts of objects compared to those distinguishing them.

Clearly such a measure can be utilized as the core of shape based image retrieval, but apart from that it offers solutions to many kinds of problems in applied computer vision. The talk will give an example how partial shape similarity offers a solution to a main problem in robotics: mapping of environment data gained by a laser range scanner.

Along with the application of robot mapping, the talk's second part will go the opposite direction (from robotics to object recognition): presenting a modified version of EM (Expectation Maximization), the modification tailoring the classical EM approach to robot mapping, the connection to EM based edge finding in images and grouping edges to object parts will be made. These parts can be the input to the aforementioned partial shape similarity measure in 2D. But the modified EM algorithm also offers interesting object decompositions into parts of 3D objects, a decomposition that offers a base for 3D similarity measures. The talk will present examples of 3D decompositions and show properties and behaviour of the algorithm to motivate possible 3D similarity measures.

Keywords: Shape similarity, object parts, robot mapping, EM

Geometric Algorithms for Transposition Invariant Content-Based Music Retrieval of Symbolically Encoded Music

Kjell Lemström (University of Helsinki, FIN)

We represent music as sets of points or sets of horizontal line segments in the Euclidean plane. Via this geometric representation we cast transposition invariant content-based music retrieval problems as ones of matching sets of points or sets of horizontal line segments in plane under translations. For finding the exact occurrences of a point set (the query pattern) of size m within another point set (representing the database) of size n , we give an algorithm with running time $O(mn)$, and for finding partial occurrences another algorithm with running time $O(mn \log m)$. We also use the total length of the overlap between the line segments of a translated query and a database (i.e., the shared time) as a quality measure of an occurrence and present an $O(n \log n + mn \log m)$ algorithm for finding translations giving the largest possible overlap.

Keywords: Content-based music retrieval, geometric algorithms

Joint work of: Ukkonen, Esko; Lemström, Kjell; Mäkinen, Veli

A User-Oriented Approach to Music Information Retrieval

Micheline Lesaffre (Ghent University, B)

Search and retrieval of specific musical content (e.g. emotion, melody) has become an important aspect of system development but only little research is user-oriented. The success of music information retrieval technology primarily depends on both assessing and meeting the needs of its users. Potential users of music information retrieval systems, however, draw upon various ways of expressing themselves. But, who are the potential users of MIR systems and how would they describe music qualities? High-level concepts contribute to the definition of meaning in music. How can we measure meaning and emotion in music? How can we define the higher-order understanding of features of music that the average users share? Information on listener's perception of qualities of music is needed to make automated access to music content attractive to system users. The emphasis of our investigation is on a user-oriented approach to the semantic description of music. We report the results of an experiment that explores how users perceive affects in music, and what structural descriptions of music best characterize their understanding of music expression. 79 potential users of music information retrieval systems rated different sets of adjectives, while they were listening to 160 pieces of real music. The subject group (79) was recruited amongst 774 participants in a large survey on the music background, habits and interests, preferred genres, taste and favourite titles of people who are willing to use interactive music systems. Moreover, the stimuli used reflected the musical taste of the average participant in the large survey (774). The study reveals that perceived qualities of music are affected by the profile of the user. Significant subject dependencies are found for age, music expertise, musicianship, broadness of taste and familiarity with classical music. Furthermore, interesting relationships are discovered between expressive and structural features. Analyses show that the targeted population most unanimously agrees on loudness and tempo, whilst less unanimity was found for timbre and articulation. Finally, our findings are tested and validated by means of a demo of a semantic music recommender system prototype that supports the querying of a music database by semantic descriptors for affect, structure and motion. The system, that recommends music from a relational database containing the quality ratings provided by the participants, illustrates the potential of a user-dependent and emotion-based retrieval of music.

Keywords: Semantic description, user studies, music information retrieval

Joint work of: Lesaffre, Micheline; Leman, Marc; Martens, Jean-Pierre

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/650>

Indexing of video sequences: specific vs interactive processing

Stéphane Marchand-Maillet (CUI - Geneva, CH)

Recent developments on video indexing have shown that this indexing is only valuable if done at a sufficient semantic level. On the other hand, it is known that such a semantic level cannot be reached without drastically constraining the domain. Hence, the indexer is then left with essentially two strategies. Either giving up on corpus genericity and therefore provide a specific automated processing of the content. Alternatively, a similar semantic level may be attained within a generic corpus at the cost of required user assistance and therefore privilege interaction.

Here, we show how we have mixed both strategies in Vicode, our video indexing platform with using salient content as a path to specificity while remaining flexible to online learning based on user relevance feedback in a QBE search paradigm.

Keywords: Video, salient content, cognition

Joint work of: Marchand-Maillet, Stéphane; Bruno, Eric ; Moenne-Loccoz, Nicolas

Full Paper:

<http://vipr.unige.ch>

Partial Matching by Structural Descriptors

Simone Marini (CNR - Genova, I)

Technological improvements related to object acquisition, visualization and modeling, have caused a considerable growth of the number of 3D models in digital form, which are now available in large databases of shapes, ranging from unstructured repositories, like the web, to specialized catalogues used in engineering and simulation. Whilst there are already techniques for rapidly extracting knowledge from massive volumes of texts (e.g., Google), it is harder to structure, filter, organize, retrieve and maintain archives of 3D digital shapes. Efficient 3D search engines are therefore necessary to avoid re-inventing (or re-designing) the wheel every time, and to support the re-use of design models.

The majority of the methods developed so far for shape matching span from coarse filters suited to browse large 3D repositories on the web, to domain-specific approaches to assess the similarity of models; they are related to four main scenarios:

1. recognition of objects whose overall shape is similar (global matching);

2. recognition of similar sub-parts in objects that are both structurally and globally similar (sub-part correspondence);
3. recognition of similar sub-parts in objects having different overall shape (partial matching);
4. recognition of a whole object as a sub-part of another (part-in-whole).

Note that the fourth case is a particular case of the third one. The majority of the methods developed so far for shape matching fit into the first category.

Most of these methods are based on geometric shape descriptors (e.g., distances, angles, curvatures, area, or volumes) which characterize the shape distribution of the object in the 3D space or visual shape characteristics. These methods are suited to solve the global matching problem, and the matching is measured by some function that provides a global value of similarity.

An important point is that the similarity between two shapes is assessed not only in terms of identical global match but it is based on the contribution of common features compared to those distinguishing them. Therefore, a retrieval system closer to the human perception of similarity should be able to assess similarity but also to identify the common parts. For example in the design applications, partial matching is definitely desirable: if an existing object differs only in a few details from a new model to be produced, then it is not difficult to adapt the existing object to the new one.

In this sense, methods for global matching do not have the capability to explicitly represent the relevant sub-parts of the object and how this sub-part are spatially related among them. Methods for sub-part correspondence and for partial matching are more elaborate in the sense that they provide not only a global measure of similarity, but also the indication of where the similarities occur in the shapes.

We believe that by differentiating the geometric and structural level of description of shapes, an automatic retrieval system will be able to provide results for sub-part correspondence and partial matching. The use of structural descriptors is an intermediate step between the geometry and the semantics of the object, since it characterizes features of the object that are relevant with respect to the application context. For example information about the shape structure may make the difference when the sub-part analysis supports decisions on how to reuse parts of a model for the design of a new one.

Since there is neither a single best shape characterization nor a single best similarity measure, we propose a framework for working on shape comparison where different characterization methods can be plugged-in and tested, while keeping the same computational setting. When the structural description can be encoded as a directed attributed graph, the sub-part correspondence between two shapes is obtained by using graph matching techniques. The proposed framework formalizes the enumeration of all common subgraphs of two graphs in a way that makes it easy to plug in heuristics, according to the specific case, thus achieving different approximations of the optimal solution.

Keywords: Partial Matching, 3D Structural Shape Descriptor, Graph Matching

Joint work of: Marini, Simone; Silvia, Biasotti; Bianca, Falcidieno

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/651>

Point-set algorithms for pattern discovery and pattern matching in music

David Meredith (Goldsmiths College - London, GB)

An algorithm that discovers the themes, motives and other perceptually significant repeated patterns in a musical work can be used, for example, in a music information retrieval system for indexing a collection of music documents so that it can be searched more rapidly. It can also be used in software tools for music analysis and composition and in a music transcription system or model of music cognition for discovering grouping structure, metrical structure and voice-leading structure.

In most approaches to pattern discovery in music, the data is assumed to be in the form of strings. However, string-based methods become inefficient when one is interested in finding highly embellished occurrences of a query pattern or searching for polyphonic patterns in polyphonic music. In our approach, these limitations are avoided by representing the music as a set of points in a multi-dimensional Euclidean space. This point-set pattern matching approach allows the maximal repeated patterns in a passage of polyphonic music to be discovered in quadratic time and all occurrences of these patterns to be found in cubic time. More recently, we have found that the best match for a query point set within a text point set of size n can be found in $O(n \log n)$ time by incorporating randomized projection, uniform hashing and FFT into our approach. Also, by using appropriate heuristics for selecting compact maximal repeated patterns with many non-overlapping occurrences, our method can be adapted for data compression. Moreover, the efficient encodings generated when this compression algorithm is run on music data seem to resemble the motivic-thematic analyses produced by human experts.

Keywords: Musical pattern discovery, point set matching, music information retrieval

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/652>

Automatic Classification and Retrieval of Motion Capture Data

Meinard Mueller (Universität Bonn, D)

The reuse of human motion capture data to create new, realistic motions by applying morphing and blending techniques has become an important issue in computer animation.

This requires the identification and extraction of logically related motions scattered within some data set. Such content-based retrieval of motion capture data, which is the topic of this talk, constitutes a difficult and time-consuming problem due to significant spatio-temporal variations between logically related motions. Recent approaches to motion retrieval apply techniques such as dynamic time warping which, however, due to their quadratic space and time complexity, are not applicable to large data sets. In our approach, we introduce various kinds of boolean features describing geometric relations between specified body points of a pose and show how these features induce a time segmentation of motion capture data streams. By incorporating spatio-temporal invariance into the geometric features and adaptive segments, we are able to adopt efficient indexing methods allowing for flexible and efficient content-based retrieval and browsing in huge motion capture databases.

Finally, a new method for automatic motion classification is presented. Using relational motion features, we introduce the concept of motion templates, by which the essence of an entire class of logically related motions can be captured.

Keywords: Motion capture, relational feature, retrieval, classification, indexing, motion template, adaptive segmentation, time alignment

Full Paper:

<http://www-mmdb.iai.uni-bonn.de/projects/mocap/>

See also: Mueller, M., Roeder, T., Clausen, M.: Efficient Content-Based Retrieval of Motion Capture Data. ACM Transactions on Graphics 24 (3), (Proceedings of ACM SIGGRAPH 2005), 677-685.

Discovery of Near-duplicate and Common Visual Concepts

Chong-Wah Ngo (City University - Hong Kong, HK)

Discovery of useful knowledge from large image and video datasets has been one of the common interests in content-based retrieval. The useful knowledge can serve as "entry points" for exploring the index structure, content summarization and semantic analysis. In this talk, we present two of our recent works in: (i) Automatic discovery of near-duplicate keyframes in broadcast videos, (ii) Automatic discovery of common pattern in images.

Near-duplicate keyframes (NDK), by definition, are similar keyframes but undergone a variations of changes including lighting, viewpoint, color, contrast and editing. While human subjects could probably identify them by spending time to browse through the entire set of keyframes, it become difficult for a machine to automatically discover them partly due to the various local variations introduced to NDKs. In this talk, we present techniques to effective detect NDK pairs and efficiently track them into groups, by utilizing both content and external cues. For content cue, we demonstrate that local interest point is a

good choice for this problem, with the appropriate choices of features, point-to-point matching strategy, and learning mechanism. For external cue, we utilize the time-span distribution and transitive nature of NDKs to rapidly grow and discover NDK groups.

A different version of the problem, compared to NDK discovery, is the automatic identification of the common patterns in images. The patterns could probably appear in background clutter under various viewpoint and scale changes. The task is potentially challenging if we simply want to find common concepts that frequently appear and potentially useful to describe an arbitrarily given image data set. For human subjects, it is easy to identify the common items of few images, but it become time consuming for a moderately large image data set. For a machine, naturally but also interestingly, it is easier to identify common items for three images than two images. To tackle this problem, we propose a novel segmentation-insensitive approach for mining common patterns from as few as two images. Based on Earth Mover's Distance (EMD), we propose an approach, namely local flow maximization (LFM), to find the best estimation of location and scale of a common pattern, under the multiple instance learning (MIL) setting. In addition to finding common pattern, We also demonstrate that this approach can yield encouraging results for image retrieval with feedback mechasim.

Content-based Retrieval in a Web 2.0 World

Jeremy Pickens (FX Palo Alto Laboratory, USA)

The classic problem of ad hoc information retrieval involves a user with an information need, a representation or expression of that information need (the query), and a system or retrieval engine that compares the query against a collection of items in order to return the most relevant items to the user information need. Despite numerous and obvious exceptions, in general text information retrieval has a fairly high correlation between the syntax of a query as expressed in language and the semantics of the information need. Textual similarity is highly correlated with relevance. On the other hand, in content-based multimedia retrieval (images, video, music, 3d models), objects encompass multitudinous semantics in many different dimensions. In music for example there are properties of pitch, tempo, rhythm, timbre, singer characteristics, genre, instrumentation, year of production, and so on. The correlation between similarity and relevance is much lower. Two music pieces might be similar because they both use similar instruments, timbres, tempos and singers, but they are not necessary both relevant to my information need if I am looking for waltzes, and one piece is in 3/4 and the other in 4/4.

The current popular solution to this problem, characterized by buzzwords such as "collective intelligence", "wisdom of crowds" and "Web 2.0", is to bypass content altogether. By instead aggregating the media interactions (playlists, tags, click behavior, etc.) of massive numbers of people, the collective intelligence

approach hopes to be able to determine relevance directly, without the need for content-based methods. If people are not only the ultimate consumers, but also the ultimate producers of relevance, why waste any effort on a problem as difficult as content-based retrieval? In our presentation we reject this notion of complete reliance on collective intelligence methods and argue that content-based methods are necessary. Aggregate crowd relevance information may be able to tell us what should be retrieved, but it still will not tell us why something was retrieved. For that, we still need to rely on the explanatory power of content.

Therefore, we propose the "cognitive disclosure" paradigm, in which semantic representations are chosen a priori by designers of a content retrieval system, i.e. content-features necessary to call a piece of music a "waltz", or to call an image a "landscape".

These semantic categories are then revealed to users at retrieval time, to allow them more intelligent selection of the types of information that is relevant to them. This problem is still very difficult and there are no easy solutions. However, our purpose is simply to explain why "wisdom of crowds" approaches will inevitably fall short, and content-based methods are still going to be necessary.

Keywords: Content retrieval, collective intelligence

Cognitive Aspects of Computing Shape

Paul Rosin (Cardiff University, GB)

An important aspect of content-based image retrieval is shape. I will survey some of the work in the area of human visual perception on shape to see what methods and models have been developed, and then follow on to give some examples of cross-over between visual perception and computer vision.

Keywords: Shape, perception, psychophysics

Visual Feature Space Analysis

Dietmar Saupe (Universität Konstanz, D)

In many important application domains such as Business and Finance, Process Monitoring, and Security, huge and quickly increasing volumes of complex data are collected and archived. Strong efforts are underway developing automatic and interactive analysis tools for mining useful information from these data repositories. At the heart of many data analysis tools is the notion of similarity which has to be defined appropriately in the given data space to allow meaningful clustering, classification, and retrieval, among other tasks. The feature vector approach is one of the most popular schemes for managing multimedia data. For many data types such as audio, images, or 3D models, an abundance of different

feature vector extractors are available. The automatic (unsupervised) identification of the best suited feature extractor for a given multimedia database is a difficult and largely unsolved problem. We address the problem of comparative unsupervised feature space analysis. We propose two approaches for the visual analysis of certain feature space characteristics contributing to estimated discrimination power provided in the respective feature spaces. We apply the approaches on a database of 3D objects represented in different feature spaces, and we experimentally show the methods to be useful (a) for unsupervised comparative estimation of discrimination power and (b) for visually analyzing important properties of the components (dimensions) of the respective feature spaces. The results of the analysis are useful for feature selection and engineering. The second approach, projection-based data space similarity visualization, is interesting (a) for visual discrimination analysis within a given similarity space, and (b) for comparative analysis of similarity characteristics of a given data set represented in different similarity spaces. We introduce an intuitive and effective novel approach for projection-based similarity visualization for interactive discrimination analysis, data exploration, and visual evaluation of metric space effectiveness. The approach is based on the hull metaphor for visually aggregating sets of points in projected space and can be used with a variety of different projection techniques. The effectiveness of the approach is demonstrated by application on two well-known data sets. We also present statistical evidence justifying the hull metaphor. We advocate the hull-based approach over the standard symbol-based approach to projection visualization, as it allows a more effective perception of similarity relationships and class distribution characteristics. This work carried out by Tobias Schreck and Christian Panse in a DFG-supported project following the V3D2 Strategic Research Initiative (Project of D. Keim and D. Saupe).

Keywords: Retrieval, feature vector, visual analysis, projection

Full Paper:

<http://dbvis.inf.uni-konstanz.de/research/projects/SimSearch3D/>

Probabilistic Algorithms for Matching Images

Ludmila Scharf (Freie Universität Berlin, D)

As a part of a content based retrieval system for figurative images we develop algorithms for determining the measure of similarity between two images. We assume that shapes contained in the images are modeled by sets of plane polygonal curves.

Our objective is to develop an algorithm which comes close to human similarity perception and allows an efficient implementation for the retrieval system.

The method we introduce attempts to capture an intuitive notion of matching, i.e., we find one or more candidates for the best transformation, that when applied to one shape maps the most similar parts of the shapes to each other.

The major idea is to take a random sample of points from both shapes and give a “vote” for that transformation (translation, rigid motion or similarity map) matching one sample with the other. If that experiment is repeated many times, we get a certain distribution of votes in the transformation space. Clusters of the votes indicate the candidates of transformations which would give the best match between the two shapes.

Keywords: Shape matching, probabilistic algorithm

Joint work of: Alt, Helmut; Scharf, Ludmila; Scholz, Sven

3D Shape Descriptions for Matching and Retrieval

Michela Spagnuolo (CNR - Genova, I)

Assessing the similarity among 3D shapes is a very complex and challenging research topic. While human perception have been widely studied and produced theories that received a large consensus, the computational aspects of 3D shape retrieval and matching have been only recently addressed.

The majority of the methods proposed in the literature mainly focus on the geometry of shapes, in the sense of considering its spatial distribution or extent in the 3D space. From a practical point of view, the main advantage of these methods is that they do not make specific assumption on the topology of the digital models, usually triangle meshes or even triangle soups. Moreover, these methods are also computationally efficient.

There is a growing consensus, however, that shapes are recognized and coded mentally in terms of relevant parts and their spatial configuration, or structure. Methods approaching the problem from a geometric point of view do not take into account the structure of the shape and generally the similarity distance between two objects depends on their spatial embedding.

The presentation will discuss the definition and use of structural descriptions for assessing shape similarity. The idea is to define a shape description framework based on results of differential topology which deal with the description of shapes by means of the properties of one, or more, real-valued functions defined over the shape. Studying these properties, several topological descriptions of the shape can be defined, which may also encode different geometric and morphological attributes that globally and locally describe the shape. Examples and results will be discussed and ongoing work outlined.

This work is partially supported by the EU Newtwork of Excellence AIMSHAPE.

Keywords: 3D shape descriptors, computational topology

Joint work of: Spagnuolo, Michela; Biasotti, Silvia; Falcidieno, Bianca; Marini, Simone

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/653>

Semantic-Oriented 3D Shape Retrieval using Relevance Feedback

Ayellet Tal (Technion - Haifa, IL)

Shape-based retrieval of 3D models has become an important challenge in computer graphics. Object similarity, however, is a subjective matter, dependent on the human viewer, since objects have semantics and are not mere geometric entities. Relevance feedback aims at addressing the subjectivity of similarity.

This talk presents a relevance feedback algorithm that is based on supervised as well as unsupervised feature extraction techniques. It also proposes a novel signature for 3D models, the sphere projection. A Web search engine that realizes the signature and the relevance feedback algorithm, is presented.

Keywords: 3D retrieval, search engine, relevance feedback

Joint work of: Leifman, George; Meir, Ron; Tal, Ayellet

Full Paper:

<http://www.ee.technion.ac.il/~ayellet/Ps/LeifmanMeirTal.pdf>

Robust Scene Categorization by Learning Image Statistics in Context

Jan Van Gemert (University of Amsterdam, NL)

We present a generic and robust approach for scene categorization. A complex scene is described by proto-concepts like vegetation, water, fire, sky etc. These proto-concepts are represented by low level features, where we use natural images statistics to compactly represent color invariant texture information by a Weibull distribution.

We introduce the notion of contextures which preserve the context of textures in a visual scene with an occurrence histogram (context) of similarities to proto-concept descriptors (texture). In contrast to a codebook approach, we use the similarity to all vocabulary elements to generalize beyond the code words.

Visual descriptors are attained by combining different types of contexts with different texture parameters. The visual scene descriptors are generalized to visual categories by training a support vector machine.

We evaluate our approach on 3 different datasets:

1. 50 categories for the TRECVID video dataset;
2. the Caltech 101-object images;
3. 89 categories being the intersection of the Corel photo stock with the Art Explosion photo stock.

Results show that our approach is robust over different datasets, while maintaining competitive performance.

Keywords: Scene Categorization, Contextual Vision, Codebook Representation, Natural Image Statistics, Color Invariance

Joint work of: Van Gemert, Jan; Geusebroek, Jan-Mark; Veenman, Cor; Snoek, Cees; Smeulders, Arnold

Properties and Performances of Shape Similarity Measures

Remco Veltkamp (Utrecht University, NL)

This paper gives an overview of shape dissimilarity measure properties, such as metric and robustness properties, and of retrieval performance measures. Fifteen shape similarity measures are shortly described and compared. Since an objective comparison of their qualities seems to be impossible, experimental comparison is needed.

The Motion Picture Expert Group (MPEG), a working group of ISO/IEC has defined the MPEG-7 standard for description and search of audio and visual content. A region based and a contour based shape similarity method are part of the standard.

The data set created by the MPEG-7 committee for evaluation of shape similarity measures offers an excellent possibility for objective experimental comparison of the existing approaches evaluated based on the retrieval rate. Their retrieval results on the MPEG-7 Core Experiment Core Experiment Shape-1 test set as reported in the literature and obtained by a reimplementation are compared and discussed. To compare the performance of similarity measures, we built the framework SIDESTEP – Shape-based Image Delivery Statistics Evaluation Project, <http://give-lab.cs.uu.nl/sidestep/>.

Keywords: Shape similarity, Evaluation

Joint work of: Veltkamp, Remco; Latecki, Longin Jan

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2006/654>

Are we Ready to Embrace the Semantic Web?

Shankar Vembu (DFKI Kaiserslautern, D)

The aim of the semantic web is to describe resources available on the web using metadata elements that can be processed or interpreted by machines. MPEG-7 is the result of a standardisation effort to annotate multimedia documents, and it offers a rich suite of metadata descriptors for describing these documents at various levels of abstraction from low-level features to high-level semantics. Owing to the proliferation of multimedia content in the internet, there is now a lot of interest in the semantic web community in multimedia metadata standards

in general, and MPEG-7 in particular. Despite the fact that the semantic web initiatives could benefit a lot from MPEG-7 for the annotation of multimedia documents, recent studies have underlined the limitations of MPEG-7 in describing the semantics of highly structured domains like sports or medicine. This has led to an upsurge of interest in adopting an integrated approach to the design of multimedia ontologies. In our work, we describe a systematic approach to the design of multimedia ontologies in which we use MPEG-7 to model only the structural and the low-level aspects of multimedia documents. High-level semantics are described using domain-specific vocabularies. A retrieval engine based on this framework will then be able to process high-level text-based semantic queries.

Whilst a lot of research has been done in the design of multimedia ontologies, a plaguing issue is the automatic annotation of multimedia content at a semantic level. It is possible to derive low-level descriptors using state-of-the-art techniques in multimedia content analysis, but the same does not hold true when it comes to analysing multimedia content at a high level of abstraction. We discuss various approaches that have been recently proposed to accomplish this task. An interesting line of discussion is the automatic population and enrichment of multimedia ontologies that offers a lot of challenges and stresses the need for efficient approaches for the semantic analysis of multimedia documents.

Keywords: Semantic analysis, Multimedia ontologies, Semantic web

Joint work of: Vembu, Shankar; Baumann, Stephan

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2006/655>

WITCHCRAFT: Melody retrieval in Dutch folksongs

Frans Wiering (Utrecht University, NL)

This talk introduces the WITCHCRAFT project, a 4-year project to design and build a usable melody search engine for folksongs. It will start on May 1st 2006, so right after this seminar. I therefore cannot present much results yet. Instead I will discuss the motivation of the project, the materials we are going to use, and some of the issues we expect to encounter.

The motivation for the WITCHCRAFT project comes to a significant extent from the cultural heritage partner, the Meertens Instituut for research and documentation of Dutch language and culture. It maintains a database of Dutch folksongs (the Liederbank) that contains a large amount of metadata. But there is no content-based melody search, and this is what the project will remedy. Specifically it will provide content-based access to a large collection of melodies known as *Onder de Groene Linde* (ca. 7000 songs). Researchers need content based searching to help them identify melodies and to trace to changes a melody underwent over time and in space. But also the general public is genuinely interested in this repertoire, so the usability of the system is an important issue.

The variability of melody, caused by oral transmission, is probably the most complex problem to deal with. To be able to do so we need to integrate quite some music cognition knowledge into the system. I will conclude my talk by performing a small experiment that illustrates the role of human memory in melody recollection.

Keywords: Music information retrieval, folksong

Joint work of: Wiering, Frans; Veltkamp, Remco C.; Grijp, Louis P.