Department of Media Technology

Journalistic Image Access: Description, Categorization and Searching

Stina Westman

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 87/2011

Journalistic Image Access: Description, Categorization and Searching

Stina Westman

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Science for public examination and debate in Auditorium AS1 at the Aalto University School of Science (Espoo, Finland) on the 21st of October 2011 at 12 noon.

Aalto University
School of Science
Department of Media Technology
Visual Media

Supervisor
Professor Pirkko Oittinen

Preliminary examiners
Assistant Professor Abebe Rorissa, University of Albany, USA
Professor Eero Sormunen, University of Tampere, Finland

Opponent
Professor Corinne Jörgensen, Florida State University, USA

Aalto University publication series
DOCTORAL DISSERTATIONS 87/2011

ISBN 978-952-60-4293-0 (pdf)
ISBN 978-952-60-4292-3 (printed)
ISSN-L 1799-4934
ISSN 1799-4942 (pdf)
ISSN 1799-4934 (printed)

Aalto Print
Helsinki 2011

Finland

The dissertation can be read at http://lib.tkk.fi/Diss/

A! **Aalto University**

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Stina Westman

Name of the doctoral dissertation
Journalistic Image Access: Description, Categorization and Searching

Publisher School of Science

Unit Department of Media Technology

Series Aalto University publication series DOCTORAL DISSERTATIONS 87/2011

Field of research Media Technology

Manuscript submitted 14 June 2011 **Manuscript revised** 9 September 2011

Date of the defence 21 October 2011 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

The quantity of digital imagery continues to grow, creating a pressing need to develop efficient methods for organizing and retrieving images. Knowledge on user behavior in image description and search is required for creating effective and satisfying searching experiences. The nature of visual information and journalistic images creates challenges in representing and matching images with user needs.

The goal of this dissertation was to understand the processes in journalistic image access (description, categorization, and searching), and the effects of contextual factors on preferred access points. These were studied using multiple data collection and analysis methods across several studies. Image attributes used to describe journalistic imagery were analyzed based on description tasks and compared to a typology developed through a meta-analysis of literature on image attributes. Journalistic image search processes and query types were analyzed through a field study and multimodal image retrieval experiment. Image categorization was studied via sorting experiments leading to a categorization model. Advances to research methods concerning search tasks and categorization procedures were implemented.

Contextual effects on image access were found related to organizational contexts, work, and search tasks, as well as publication context. Image retrieval in a journalistic work context was contextual at the level of image needs and search process. While text queries, together with browsing, remained the key access mode to journalistic imagery, participants also used visual access modes in the experiment, constructing multimodal queries. Assigned search task type and searcher expertise had an effect on query modes utilized. Journalistic images were mostly described and queried for on the semantic level but also syntactic attributes were used. Constraining the description led to more abstract descriptions. Image similarity was evaluated mainly based on generic semantics. However, functionally oriented categories were also constructed, especially by domain experts. Availability of page context promoted thematic rather than object-based categorization.

The findings increase our understanding of user behavior in image description, categorization, and searching, as well as have implications for future solutions in journalistic image access. The contexts of image production, use, and search merit more interest in research as these could be leveraged for supporting annotation and retrieval. Multiple access points should be created for journalistic images based on image content and function. Support for multimodal query formulation should also be offered. The contributions of this dissertation may be used to create evaluation criteria for journalistic image access systems.

Keywords image journalism, image attributes, image search

ISBN (printed) 978-952-60-4292-3 **ISBN (pdf)** 978-952-60-4293-0

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2011

Pages 166 **The dissertation can be read at** http://lib.tkk.fi/Diss/

Aalto-yliopisto

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tiivistelmä

Tekijä
Stina Westman

Väitöskirjan nimi
Journalistinen kuvatiedonhaku: kuvailu, luokittelu ja haku

Julkaisija Perustieteiden korkeakoulu

Yksikkö Mediatekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 87/2011

Tutkimusala Mediatekniikka

Käsikirjoituksen pvm 14.06.2011 **Korjatun käsikirjoituksen pvm** 09.09.2011

Väitöspäivä 21.10.2011 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Digitaalisten kuvajoukkojen jatkuva kasvu luo paineita kehittää tehokkaampia tapoja järjestää ja hakea kuvia. Toimivien ja tyydyttävien hakukokemusten tuottamiseen tarvitaan tietoa kuvailu- ja hakukäyttäytymisestä. Visuaalisen tiedon ja journalististen kuvien luonne tekee kuvien esittämisestä ja sovittamisesta käyttäjätarpeisiin haastavaa.

Tämän väitöskirjan tavoitteena oli ymmärtää journalistisen kuvahaun prosesseja sekä kontekstuaalisten tekijöiden vaikutuksia preferoituihin kuvailutasoihin ja –attribuutteihin. Näitä selvitettiin kuvailun, luokittelun ja haun osalta käyttäen useita aineistonkeräys- ja analyysimenetelmiä useassa eri tutkimuksessa. Journalististen kuvien kuvailussa käytettäviä attribuutteja analysoitiin kuvailutehtävien perusteella ja niitä verrattiin kirjallisuuden pohjalta laadittuun typologiaan. Journalistisia kuvahakuprosesseja ja kyselytyyppejä analysoitiin kenttätutkimuksen ja multimodaalisen kuvahakukokeen avulla. Kuvien luokittelua tutkittiin pohjautuen lajittelukokeisiin ja kuville laadittiin luokittelumalli. Tutkimusmenetelmiä kehitettiin edelleen kuvahakutehtävien ja lajitteluproseduurien osalta.

Kontekstuaalisia vaikutuksia kuvahakuun oli organisatorisella kontekstilla, työ- ja hakutehtävillä sekä kuvien julkaisukontekstilla. Journalistinen kuvatiedonhaku oli kontekstiriippuvaista sekä kuvatarpeiden että hakuprosessien osalta. Vaikka tekstihaku – selailun ohella – oli edelleen tärkein hakutapa journalistisille kuville, kokeen osallistujat käyttivät myös visuaalisia hakutapoja. Hakutehtävätyyppi ja hakijan asiantuntemus vaikuttivat siihen, mitä hakutapoja multimodaalisessa kuvahaussa käytettiin. Journalististen kuvien kuvailu ja haku tapahtui enimmäkseen semanttisella tasolla, mutta myös syntaktisia attribuutteja käytettiin. Kuvailun rajoittaminen johti abstraktimpaan kuvailuun. Kuvien samankaltaisuutta arvioitiin enimmäkseen geneerisen semantiikan tasolla. Varsinkin asiantuntijat muodostivat kuitenkin myös funktionaalisesti suuntautuneita kategorioita. Julkaiskontekstin näkyminen johti temaattiseen kuvaluokitteluun kohdepohjaisen luokittelun sijaan.

Tulokset lisäävät ymmärrystämme kuvailu-, luokittelu- ja hakukäyttäytymisestä ja niillä on implikaatioita tuleviin ratkaisuihin journalistisessa kuvahaussa. Kuvien tuotannon, käytön ja haun kontekstia tulisi tutkia edelleen jotta sitä voitaisiin hyödyntää kuvailun ja haun tukemiseen. Kuvahaussa tulisi tarjota useita hakutapoja perustuen kuvien sisältöön ja funktioon. Multimodaalista hakua tulisi myös tukea. Tämän väitöskirjan kontribuutioita voidaan käyttää hyväksi laadittaessa arviointikriteerejä journalistisille kuvahakujärjestelmille.

Avainsanat kuvajournalismi, kuva-attribuutit, kuvahaku

ISBN (painettu) 978-952-60-4292-3 **ISBN (pdf)** 978-952-60-4293-0

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2011

Sivumäärä 166 **Luettavissa verkossa osoitteessa** http://lib.tkk.fi/Diss/

Preface

The research for this dissertation was carried out during the years 2005-2011 in the Visual Media research group at the Department of Media Technology (previously Laboratory of Media Technology) at Aalto University (previously Helsinki University of Technology), in Finland.

Funding for the research has been provided through the project Visual Commons in the Motive program of the Academy of Finland, via several projects funded by the Finnish Funding Agency for Technology and Innovation, and grants by the Graphic Industry Research Foundation, Patricia Seppälä Foundation, as well as Jenny and Antti Wihuri Foundation.

I am grateful to my supervisor, Professor Pirkko Oittinen, for introducing me to the research field of imaging, and giving me her time, feedback, and support over the years. I also wish to acknowledge the valuable mentoring by Professor Emeritus Hannu Saarelma throughout my studies. I value the comments and suggestions for improvement by Assistant Professor Abebe Rorissa and Professor Eero Sormunen, pre-examiners of this dissertation.

This work would not have been possible without the help, encouragement and support from several people and organizations. I wish to thank all organizations and individuals who participated in the data collection activities in the studies in this dissertation. I am especially indebted to Anna Suominen (formerly at Lehtikuva), Irmeli Toivanen and Kari Rouhiainen (Keskisuomalainen) as well as Maija Töyry (Aalto University School of Art and Design).

Several colleagues have contributed to this dissertation. Especially I would like to thank my coauthor and comadre Mari Laine-Hernandez, always willing to share intellectual debates over baked goods. I would also like to acknowledge the valuable work by my bright Master's thesis students Anna Berg, Antti Lustila and Saara Sulanto. Gratitude is due towards Raisa Halonen for enduring me in a shared office space and solving the mystery of my disappearing figures.

Finally I want to thank my family and friends for their continuing support and encouragement. Particular thanks to Mikko for nourishing me during days of work and leisure.

Espoo, 9st of September, 2011

Stina Westman

Contents

List of Publications and Author's Contribution

This dissertation is based on the following six publications. They are referred to in the dissertation by their Roman numerals (I-VI).

I Laine-Hernandez, M., & Westman, S. (2006). Image Semantics in the Description and Categorization of Journalistic Photographs. *Proceedings of the American Society for Information Science and Technology, 43*(1):1–25.

II Westman, S. (2009). Image Users' Needs and Searching Behaviour. In A. Göker, & J. Davies (Eds.), *Information Retrieval: Searching in the 21st Century* (pp. 63-84). Chichester: John Wiley & Sons.

III Westman, S., & Oittinen, P. (2006). Image Retrieval by End-users and Intermediaries in a Journalistic Work Context. In I. Ruthven, P. Borlund, P. Ingwersen, N.J. Belkin, A. Tombros, & P. Vakkari (Eds.), *Proceedings of the First Symposium on Information Interaction in Context (IIiX)* (pp. 102-110). New York: ACM.

IV Westman, S., Lustila, A., & Oittinen, P. (2008). Search Strategies in Multimodal Image Retrieval. In P. Borlund, J.W. Schneider, M. Lalmas, A. Tombros, J.F. Loughborough, D. Kelly, & A.P. de Vries, (Eds.), *Proceedings of the Second Symposium on Information Interaction in Context (IIiX '08)* (pp. 13-20). New York: ACM.

V Westman, S., & Laine-Hernandez, M. (2008). The Effect of Page Context on Magazine Image Categorization. *Proceedings of the American Society for Information Science and Technology, 45*(1):1-11.

VI Westman, S., Laine-Hernandez, M., & Oittinen, P. (2011). Development and Evaluation of a Multifaceted Magazine Image Categorization Model. *Journal of the American Society for Information Science and Technology, 62*(2): 295-313.

Publication I was an equal collaboration with names given in alphabetical order. The author was in charge of the design, analysis and reporting of the keywording task in the experiment conducted. The author was the sole contributor in publication II. The author was the main contributor in publications III-VI. She was responsible for designing the studies, collecting and analyzing the data and reporting the results while coauthors have contributed to the design of the user experiments (V-VI), participated in data collection activities (IV-VI) and offered comments on manuscripts (III-VI).

1 Introduction

The rapid and continuing growth in the quantity of visual content, digital images and video, creates pressing needs for more effective methods for organizing, searching, and retrieving this content. The development of image indexing and retrieval techniques has received increased attention in recent years. Knowledge concerning user behavior and preferences is also needed to create effective and satisfying search experiences. The specific nature of visual information creates challenges in representing and matching information objects and user needs.

Information access refers to the continuum of information activities from information needs to the search process, corresponding retrieval techniques, and information use. This indicates a shift from a basic information retrieval approach toward users and their behavior (Agosti, 2007). Information access is also used to describe an area of research at the intersection of computer, information, and library sciences. Information access approaches contain variety of involved entities: the content, the people involved and the context of use (Chalmers, 1999).

The focus of this work is on journalistic image access. It addresses questions related to the description, categorization and retrieval of visual materials in the context of image journalism and its tasks. Image description, categorization and retrieval are connected to research in computer, information and library science, and more specifically to the fields of image indexing and classification, as well as concept- and content-based image retrieval. As a whole, this work addresses issues relevant to digital imaging, media asset management and information retrieval in context.

1.1 Motivation and Background

Information retrieval is dependent on information representation and organization via the processes of cataloging, indexing and classification (Rowley & Hartley, 2008). Descriptive cataloging records information on the external features of information resources (e.g., title, creator) while the indexing and classification processes focus on describing the intellectual subjects of works. Indexing refers to creating representations of information objects, especially via access points. Access points are names, words or other elements that may be utilized to obtain these objects from a retrieval system. Indexing may be based on controlled vocabularies or natural language terms, sometimes called free terms (Lancaster, 1998). Alternative subject descriptions support the representation of a document through different access points.

Classification is the process of assigning information resources to classes according to a set of predetermined principles, and may be thought of as a special case of indexing. The term categorization is often used synonymously with classification (Taylor & Joudrey, 2009). However, categorization may also be defined as the cognitive process of drawing associations between documents based on simple similarities while classification is based on a set of predetermined principles (Jacob, 2004). In computer science, the term classification carries the specific notation of an algorithmic procedure for assigning a given piece of input data into one of a given number of categories (Bishop, 2006).

Images may be considered to be a special case for subject description because of the distinct nature of visual information (Chen & Rasmussen, 1999; Jespersen & Jespersen, 2004). Information in the visual modality cannot be broken down into units of meaning like text; thus interpretation of images is inherently contextual and subjective. Descriptions of images have been achieved by means of manual, intellectual selection of index terms (concept-based or human-based indexing) and automated analysis of image features (content-based or automated indexing) (Jörgensen, 2001; Rasmussen, 1997). Intellectual indexing is time-consuming; therefore, the need for automatic subject analysis of image content is evident. However, automatic indexing has its limitations because the analysis of low-level syntactic features often does not match human interpretations of image semantics leaving the so-called semantic gap (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

The development of content-based image retrieval has brought about many new techniques and systems in the last decade (Datta, Joshi, Li, & Wang, 2008). To build better image retrieval solutions there is also a need for a human-centered approach to image access, creating an understanding of factors that affect the way images are searched (Jaimes, 2006). It is necessary to understand how users interact with image retrieval systems and what types of image attributes are utilized to gain access to images. Knowledge of the types of image queries searchers construct and the search strategies they choose coupled with information on their ability and willingness to use different image access modes could inform the development of retrieval algorithms and their applications.

Regardless of the approach taken for subject description, the question of what to index remains when supporting end-user searching. The range and type of attributes needed for describing image content are still under debate. Images may be analyzed and described at various levels (Shatford, 1986). Image categorization is a central issue in image retrieval because of the need to represent and offer access not only to individual images but also to meaningful groupings of images. Subject access concerns itself with concepts and their labels as indexing terms as well as the relationships between these concepts in classification structures (Rowley & Hartley, 2008). Categorization is a way to index groups of images collectively, thereby moving some of the onus of image understanding from the point of retrieval to the point of indexing (Shatford Layne, 1994). In retrieval systems, images may be grouped to support, for example, browsing and exploratory searches. Research has aimed to discover the criteria that people apply when evaluating image similarity and the types of descriptions that people give to individual images and groups of images. Knowledge of image descriptions and similarity evaluations may be used to develop ontologies for image indexing and to evaluate algorithmic approaches to image classification.

The media industry has launched various activities aimed at enhancing the subject description of visual materials for the purposes of media asset management. The furthest developed image exchange standard is the IPTC Photo Metadata (IPTC, 2008a) which aims to describe photographs and administer them, providing the most relevant rights information. The standard includes fields for a description (e.g., "who, what, why"), keywords, scene code (e.g., night scene) and subject code (e.g., politics) (IPTC, 2008b). The subject codes were originally created for article text and were later applied to visual media. To address the special challenges of describing images, IPTC is currently developing a controlled vocabulary, which is aimed at interoperability in the image business. The vocabulary consists of terms for objects, named entities, broad topics, activities, abstract concepts, as well as conceptual and visual descriptors (Saunders, 2010). These efforts reflect the current need for descriptive attributes of images in the media.

Contextualized work tasks in specific domains and situations give rise to certain types of information seeking activities. Context in information retrieval has been conceptualized by various researchers (Ingwersen & Järvelin, 2005; Saracevic, 1997; Sonnenwald, 1999; Wilson, 1999). Context may be analyzed on different dimensions (Cool & Spink, 2002): the information environment level (e.g., work tasks), the information seeking level (e.g goals of information retrieval), the information retrieval interaction level (e.g., search strategies) and the query level. Context affects the treatment of information objects, such as images, on these multiple levels. The context and tasks that give rise to image search and the images' intended future uses affect image descriptions and queries.

Within this dissertation the term journalistic image refers to images (e.g., photographs) taken or searched for the purpose of professional publishing, including, but not limited to, newspapers and magazines. Journalistic imagery is limited by photographic conventions and style as photographs are typified through the photojournalistic processes through which they are created and published (Hall, 1981). Journalistic image collections are dynamic as images are uploaded daily from various sources. The attached metadata and descriptions may vary in structure and in quality. Journalistic image collections include a wide topical range of images that are published in multimodal contexts. These characteristics of the domain, among others, may bring about specific behavior in image description, categorization, and searching.

1.2 Contributions

This dissertation builds upon field studies and user experiments in journalistic image description, categorization, and searching. These studies aim at furthering the understanding of user behaviour in image interaction tasks and contextual effects on preferred access points to journalistic images. Contextual effects are accounted for by organizational contexts, work and search tasks and publication context of images. These factors span all the context levels defined by Cool and Spink (2002). Among these results are the first results on the effect of domain experience (VI) and page context (V) on image categorization behavior. Contributions are offered in:

- *Image description*: Empirical results on the attributes prevalent in the descriptions of journalistic images are presented (I). The effects of the description task and constraining image description are also discussed (I). The work serves to suggest access points to journalistic imagery and develop indexing approaches. A typology of image attributes based on the literature is offered in (II).
- *Image searching behavior*: Common image access strategies in journalistic image retrieval are discussed. These include a model of the journalistic image search and selection process and types of image needs expressed in work tasks through requests and queries (III), as well as tactics in multimodal image search (IV).
- *Image categorization:* A magazine image categorization model (VI) is developed based on user-supplied categories (V-VI). The categorization criteria are confirmed by a second study (VI). Dimensions of image similarity evaluations are shown (VI). The categorization model may be used to develop categorization practices and to enable image retrieval and reuse by the editorial staff. The final model is discussed in relation to current standards and the state-of-the-art in automated image classification. This work serves as a meet-in-the-middle requirements analysis for ontology development for magazine image indexing.
- *Methods in image interaction studies:* Intertwined with reported empirical results are improvements to research methods. Within the study on multimodal image searching behavior a test set of image search tasks is introduced (IV). A common categorization procedure is extended to allow multifaceted categorization and to evaluate the connections between categories (VI).

As a whole this dissertation continues and extends the research track established by Ørnager (1995, 1997) who studied the indexing of journalistic images, and by Markkula and Sormunen (1998, 2000) who contrasted indexing practices of archivists at a newspaper with image searches by journalists. The results from Ørnager are complimented by an in-depth study of magazine image categorization (V-VI) and a confirmation of image searcher types in newspapers (III). The results on journalistic image searching behavior from Markkula and Sormunen are verified and contrasted (III). Also, content-based image access modes are evaluated in addition to concept-based approaches (IV).

1.3 Structure of the Dissertation

In the following sections previous research is reviewed, followed by the description of the research methods and experiments of the studies completed for this dissertation. Next, the results are presented and discussed, followed by the conclusions.

Section 1 discusses related research in the areas of image attributes, image categorization, and image search strategies that form the body of this dissertation. Section 2 concludes with a review of previous work in journalistic image access. Section 3 provides an overview of the research approach and reports on the methods and data in the studies. Section 4 presents and discusses the results for four thematic areas: image search processes, queries and requests, image descriptions and image categorization. There is also a discussion of contributions across the studies. Section 5 presents the conclusions drawn.

2 Related Research

Images may be described and categorized, requested from intermediaries and queried by end-users on different semantic and syntactic levels. Image attributes have been typified based on theoretical analysis and user-supplied image descriptions. Previous research on image searching behavior has largely involved analyzing the content of requests and queries. Studies may focus on sessions, searches, queries or query terms (Jörgensen & Jörgensen, 2005). A search session is a series of transactions (e.g., queries, views, downloads) aimed at retrieving one or more images. A search is a set of related queries within a certain time frame that address the same topic. A query is one or more query terms in a single interaction that results in a system display of the retrieval set. Most often the analysis has been on the level of queries or query terms. Case studies have also addressed searches while transaction logs have enabled the analysis of sessions.

2.1 Image Access Points

Because of the semantic gap and the resulting need to employ intellectual descriptions of images the image indexing process is likely to remain dependent on linguistic and conceptual inferences about images. Studies have been conducted to investigate how people interpret and describe image content. Research has also been aimed to discover image categorization behavior and criteria applied when evaluating similarities between images. Based on these studies the labels assigned to images and groups of images have been analyzed for image attributes.

2.1.1 Image Attributes

Images have various attributes that may be important to users. Various conceptual frameworks built on the elements or properties of images illustrate this point. Some theories structure images as such (Burford, Briggs, & Eakins, 2003; Eakins, Briggs, & Burford, 2004) while others classify descriptions of images (Hollink, Schreiber, Wielinga, & Worring, 2004; Jörgensen, 1998), and some are meant to be used in the indexing of images (Jaimes & Chang, 2000). Nevertheless, they all classify perceived and interpreted characteristics of image content.

Jaimes and Chang (2000) developed a conceptual model for describing visual content. This model relies in its classification on the amount of knowledge required to identify and index attributes on each level. The first four levels are perceptual levels on which no world knowledge is needed (Type/technique, Global distribution, Local structure, and Global composition). The six remaining levels are conceptual levels (Generic object or scene, Specific object or scene, and Abstract object or scene). General, specific or abstract world knowledge is required to formulate image descriptions on these levels. Additionally, Jaimes and Chang also present three classes of nonvisual image information: Biographical information, Associated information and Physical attributes. Burford et al. (2003) and Eakins et al. (2004) proposed a taxonomy of image content based on a survey of literature on computer science, art history and psychology. They list ten categories of information associated with an image: Metadata, Perceptual primitives, Geometric primitives, Visual relationships, Visual extension, and Semantic units as well as Contextual, Cultural, Technical, and Emotional abstractions. The last nine categories are thought to be roughly hierarchical and to reflect the way that meaning is constructed from images.

Jörgensen (1998) analyzed user behavior in different image description tasks. She presented 12 classes of image attributes used by the participants: Object, People, Colour, Visual elements, Location, Description, People-related attributes, Art historical information, Content/story, Abstract concepts, External relationships and Viewer response. Jörgensen further classified these attribute classes into the perceptual, interpretational and reactive. Hollink et al. (2004) presented a three-level framework for the classification of image descriptions: nonvisual, perceptual and conceptual. Following Shatford (1986) and Jaimes and Chang (2000) they also divided the conceptual level, which involved image semantics, into three sublevels for general, specific and abstract image descriptions. For a complete description of an image, these levels may all be used at once.

Jörgensen's results show that people mostly use interpretational attributes related to objects, people and story as well as other semantic terms when describing image content. Similarly, Hollink et al. determined that the conceptual level was the main level of image descriptions both for free description and querying. Results from Rorissa (2008) indicate that images are described more on the basic, concrete level than on the abstract or specific levels of concepts. Greisdorf and O'Connor (2002a) showed that, in addition to generic semantic attributes, images may also be described with affective and syntactic terms. When image attributes are meant to function as indexing keywords, interindexer consistency has been evaluated (Markey, 1984; Beaudoin, 2008). Findings indicate there is a rather low consistency between indexers. This result may occur because of the inherent complexity of visual information objects, which requires more individual interpretation, thereby leading to more variation.

In addition to attributes of image content, the functions of images (i.e. their purposes of use) are of central importance in image retrieval, because the intended use of the image affects the search strategies (Jaimes, 2006). Image functions should also be considered from the viewpoint of image categorization, because the purpose of use is a possible categorization criterion. Image functions have been reviewed based on, for example, studies on image searching behavior (Conniss, Ashford & Graham, 2000) as well as informational and pedagogical functions of images in professional use (Pettersson, 1998). Images may have several types of functions, such as illustrative, informative, persuasive, attention-related (e.g., attracting and holding attention), aesthetic (e.g., adorning something) and affective (e.g., establishing a mood). Further work has been conducted for example on the effect of work tasks on the use of images as illustration or information (McCay-Peet & Toms, 2009).

2.1.2 Image Categorization and Similarity

Categorization is a basic cognitive function that is aimed at organizing groups of information objects based on similarity. Sorting techniques may be utilized to elicit the categorization structures utilized by people. Prior research has discovered that people evaluate image similarity primarily based on the presence of people and the type of scene depicted in the photographs. They distinguish, for example, between people, animals and inanimate objects as well as between whether the scenes and objects in the images are man-made or natural, such as, urban scenes vs. landscapes (Rogowitz, Frese, Smith, Bouman, & Kalin, 1998; Teeselink, Blommaert, & de Ridder, 2000). Photographs of people may be further categorized based on their gender, pose, actions and facial expressions (Rorissa & Hastings, 2004). People mostly evaluate the similarity of images at a conceptual level constructing semantic image categories. This is also the level mainly employed in other image description tasks as previously discussed.

However, people do not always form generic semantic image categories. Categories are also created based on abstract concepts related to emotions or atmosphere, cultural references and visual elements (Greisdorf & O'Connor, 2002b; Jörgensen, 1995). Sormunen, Markkula and Järvelin (1999) showed that journalists evaluate image similarity based on a variety of criteria ranging from syntactic (e.g., shooting distance, composition) to abstract (e.g., facial expression, theme).

When constructing image categories, people seem to evaluate image similarity by considering overall similarity across all of the dimensions rather than the maximal similarity along one dimension (Greisdorf & O'Connor, 2002b). This strategy results in categorizations based on multiple simultaneous similarity criteria. Jörgensen (1995) found that one third of image group names provided by participants consisted of multiple words, of which one modified the main attribute. Mojsilovic and Rogowitz (2002a) also discuss the tendency of participants to use category names to determine links between categories. When participants are allowed to place an image into multiple categories of their own previous construction, there can be considerable overlap in categorization (Laine-Hernandez & Westman, 2008). These results reflect the multifaceted nature of image description and categorization.

Category names have been analyzed as descriptions of image groups. Jörgensen (1995) showed that names assigned to groups of illustrations often included references to abstract concepts and themes, story attributes, and art historical attributes, including type and style. Names consisting of multiple terms most often referred to people and objects. Basic level theory has also been employed to show that image category labels are mostly on the basic (e.g., chair) or superordinate (e.g., furniture) level (Rorissa & Iyer, 2008). An emerging stream of research has focused on user-driven image grouping and tagging in online collections. Groups of images shared across multiple users are more often labeled with thematic terms and abstract concepts than personal photosets (Stvilia & Jörgensen, 2009). In a similar vein, groups of images have more often superordinate labels while single image descriptions are mainly at the basic level (Rorissa, 2008). The types of image tags used, for example, on Flickr may often not be found in federated image collections (Rorissa, 2010) or indexing thesauri (Stvilia & Jörgensen, 2010), creating a disconnect between users' terms and indexing terms.

2.2 Image Access Strategies

A search strategy may be thought of as a plan that the searcher has for completing the search (Bates, 1979). Search strategies consist of combinations of search tactics, i.e., moves (Fidel, 1985), which are made to further a search. Within image searches various search tactics may be distinguished: different types of queries, query reformulations, browsing, comparisons, and relevance assessments. Search patterns may be analyzed based on repeating structures of search tactics. Different search strategies seem to be preferred for different types of image needs and search tasks.

Images may be retrieved through various access modes: textual queries, visual queries, or a combination of the two. They may also be accessed by browsing structured or unstructured collections. Image needs may be formalized into queries input into digitized image collections directly by the searcher, or into written or verbal requests to intermediaries. Several studies have categorized the content of image queries or requests within specific domains such as media (Enser, 1993; Markkula & Sormunen, 2000), art history (Choi & Rasmussen, 2003; Hastings, 1995), advertising and design (Jörgensen & Jörgensen, 2005), medicine (Keister, 1994), and various other professional and academic areas (Armitage & Enser, 1997) as well as web searching (Cunningham & Masoodian, 2006; Goodrum & Spink, 2001; Jansen, 2008).

2.2.1 Image Search Process

Information retrieval is an iterative process in which searchers pick up documents and information throughout the course of a search (Bates, 1989). Conducting an image search has also been shown to be iterative (Garber & Grunes, 1992; Hastings, 1995). Searchers gain new knowledge of their retrieval task and the image collection through the search process and the image sets retrieved (Greisdorf & O'Connor, 2002a). This scenario means that their perception of the search task and the image selection criteria is dynamic.

An image search process may be thought to consist of several phases (Conniss et al., 2000). In the starting phase the searcher identifies an image need or receives an image request. The user then develops criteria for assessing the suitability of the images found, forming a mental model of the target image (Frost, 2001). The intended use of the image provides a context for the search and is the basis for creating relevance criteria. For example, the relevance judgments of journalists depend on their work task and related contextual factors such as the article to be illustrated, the page layout and the illustration styles of the newspaper (Markkula & Sormunen, 2000). Consideration of the context of use also covers such factors as image overuse, intended audience, and copyright (Conniss et al., 2000).

During the scoping and applying phases the searcher decides on a search strategy and implements it. The availability of image collections, time and financial constraints, and the need for involving an intermediary all affect the decisions of how to search. Image users attempt to use sophisticated search strategies but are often unable to apply the methods correctly (Jörgensen & Jörgensen, 2005). The level of domain knowledge that the searcher possesses has an effect on the search strategies and effectiveness (Frost et al., 2000; Matusiak, 2006). The experience that users have with image retrieval systems affects their searches in at least two ways (Eakins & Graham, 1999): an expert user is able to use suitable advanced search features but also the query terms are affected as the user gains knowledge on how retrieval systems work.

Searching may involve modifications to the first query as iteration to the search. Most image queries are modified after the initial query (Goodrum, Bejune, & Siochi, 2003; Jörgensen & Jörgensen, 2005), more often so than text queries (Goodrum & Spink, 2001). Various types of query modifications exist at the query level as terms may be added, deleted, substituted, narrowed, broadened or reintroduced (Jörgensen & Jörgensen, 2005). Query modifications may also result from changes in the search task or new retrieval ideas generated during the search. When searchers are provided with a workspace they create different result sets that reflect different semantic facets of the search task (Urban & Jose, 2006). Hung (2005) describes searchers changing query terms to match the content or textual description of the images that result from the initial search. Jörgensen and Jörgensen (2005) determined that users employed terms from image captions to modify their queries. On a related note, Conniss et al. (2000) describe searchers using the categorization of images to learn more about the subject area.

Selecting images may occur as a single action or in stages by first selecting a subset of relevant images and then selecting the image(s) to be used. During the search process and when deciding that a search has been completed, the relevance of images must be evaluated. Relevance refers to the match of the image to the image need and it may be judged on multiple levels. Topicality functions often as the first relevance criterion employed during the search process and may be the single most important criterion (Markkula & Sormunen, 2000; Choi & Rasmussen, 2002). Beyond topicality, image selection is based on both compositional and informational criteria including, for example, quality, visual appearance, and physical size (Conniss et al., 2000). The textual descriptions associated with images are an important source of information when judging relevance (Choi & Rasmussen, 2002; Hung 2005; Markkula & Sormunen, 2000). The final criteria are subjective and affective such as aesthetic criteria and emotional reactions to images (Choi & Rasmussen, 2002; Conniss et al., 2000; Markkula & Sormunen, 2000).

2.2.2 Conceptual and Visual Search Strategies

The conceptual and visual search strategies employed in image searches create search patterns. Most interactions in image search involve queries followed by viewing images or surrogates (Goodrum et al., 2003) as well as browsing and enlarging images (Hung, 2005). Jörgensen and Jörgensen (2005) determined that image queries were frequently modified but often ultimately unsuccessful. The level of complexity in the image query and search affects the usefulness of access points, and whether textual or visual access is required (Hastings, 1995).

Different types of search topics lead to different types of search tactics. General search topics easily lead to multiple queries and heavy browsing while specific needs are more likely to result in fewer queries and shorter browsing sessions (Hung, 2005; Markkula & Sormunen, 2000). Conceptual and abstract image needs seem to lend themselves more naturally to browsing rather than querying (Conniss et al., 2000), probably because they may be difficult to translate into query terms. Overall, well-defined search tasks encourage direct queries while less concrete tasks make users prefer browsing (Frost et al., 2000). When abstract images are queried, the search leads to a high number of query reformulations and iteration (Hung, 2005).

Textual queries for images tend to be short, consisting of a single term or phrase. Web image queries typically contain between two and three terms (Choi, 2009; Cunninghan & Masoodian, 2006; Goodrum & Spink, 2001; Goodrum et al., 2003; Tjondronegoro, Spink, & Jansen, 2009), whereas specialist queries in closed collections include slightly less than two terms, on average (Jörgensen & Jörgensen, 2005). This apparent shortness of the queries does not contradict the fact that the image requirements may be complex because some needs may be difficult to verbalize. The query terms vary, and the most frequent query terms are present only in a few percent of the queries (Goodrum & Spink, 2001; Jörgensen & Jörgensen, 2005). There are fewer empirical results on image search sessions than on queries due to challenges in session identification. Image search sessions typically last between five and 20 minutes and include on average two to three queries (André, Cutrell, Tan, & Smith, 2009; Goodrum & Spink, 2001; Jörgensen & Jörgensen, 2005; Tjondronegoro et al., 2009).

An important visual access mode to images is browsing. Browsing allows users to recognize images that interest them, rather than needing to formulate a precise query, and allows them to discover images of which they were previously unaware (Frost et al., 2000). As a result, browsing is an attractive search strategy, especially for users with little prior knowledge on the domain or collection. During browsing, images may be viewed and evaluated within their own medium of expression which is useful because visual impressions may be difficult to communicate verbally (Heidorn, 1999). Jörgensen and Jörgensen (2005) stress the importance of a useful browsing interface, especially during the final image selection process. Search sessions, overall, include a considerable amount of viewing result images, and comparing them (Hung, 2005). For example, Goodrum et al. (2003) report that in image searches on the web, two-thirds of the time is spent on browsing. Users seem to like browsing because of the control it gives them, and they browse the most when searching for abstract or subjective topics (Markkula & Sormunen, 2000). This observation suggests that users adopt a browsing strategy when they are unable to formulate a query. Browsing thus compensates for the difficulties of a nonexpert user in forming a query and is most useful when the user has ill-defined goals for the task (McDonald & Tait, 2003). However, browsing has been found to benefit also expert searchers (Frost et al., 2000). A key issue for browsing is that the image groups browsed must make sense to the user (Barnard & Forsyth, 2001; McDonald & Tait, 2003). For this reason, providing a meaningful categorization as an access point to image collections is important.

The search strategies that users adopt in content-based image retrieval have also received some attention. Image searchers may issue visual queries, specifying local or global levels of syntactic features or the spatial location of image elements (Eakins, 1996). In a survey by Eakins et al. (2004) visual query modes were thought to be less important than conceptual access to images. Color was an important visual query mode for those respondents who used it, and sharpness was considered important by all. Frost (2001) explored the use of visual similarity searches that were generated based on seed images found through text queries. She determined that these content-based search features were useful in sparking search ideas and were preferred by users who sought graphically appealing and eye-catching images. McDonald and Tait (2003) investigated searchers' abilities to construct visual queries with sketch and structured browsing tools. The sketch tool was useful in searches for previously seen images. Object and abstract images were primarily browsed by users. McDonald and Tait concluded that a sketch-based tool might be more appropriate for simple images, such as landscapes, because sketching requires a priori knowledge of the image composition and colors. Overall, different query input methods seem to be preferred for different types of image searches. Eakins et al. (2004) found that users prefer to type in search terms rather than to select terms from a list. However, they expressed a preference for selecting content-based search elements such as shape or texture from a menu, rather than entering them freehand.

2.2.3 Types of Image Queries and Requests

Images are most often queried or requested based on image semantics, in other words, concepts interpreted from images. Jörgensen (1999) presented a typology of image requests based on literature, including requests for the following: 1) a specific image, 2) a specific instance of a general category of images (e.g., a named person, group, thing, event, location or action), 3) a general topical or subject category of images (e.g., a type of person, group, thing, event, location or action), and 4) images communicating a specific abstract concept or affective response (e.g., warmth). Fidel (1997) suggested that image retrieval tasks may be mapped onto a continuum between data and object poles. At the data pole, images are used as information sources. The images are thus retrieved based on them containing certain data. At the objects pole images are needed as objects.

In a seminal work, Enser (1993) employed two dimensions in the coding of image requests: nonunique/unique and nonrefined/refined. Nonunique requests contained generic concepts (e.g., "bridge") while unique requests asked for specific objects, locations or events (e.g., "London Bridge"). Both classes may be refined with specifications of time, location, event or technical attributes. Most requests (69%) in Enser's sample belonged to the unique category and were refined (52%). The importance of specific instances was pronounced in the requests that originated from newspaper and magazine publishers.

Armitage and Enser (1997) extended Enser's work by employing the Panofsky-Shatford facet matrix (Table 1). The matrix is based on Shatford's (1986) refinement of Panofsky's theory of the three levels of meaning in images. The generic and specific levels cover factual content of the image and correspond to Enser's (1993) nonunique and unique levels, respectively. The third analysis level covers abstract, expressional content. Shatford added four facets (who/what/where/when) to each level creating an analytical matrix structure.

Table 1. Panofsky-Shatford facet matrix (Armitage & Enser 1997)

Facet	**Iconography (Specifics)**	**Generic, pre-iconography (Generics)**	**Iconology (Abstracts)**
Who?	Individually named person, group, thing S1	Kind of person or thing G1	Mythical or fictitious being A1
What?	Individually named event, action S2	Kind of event, action G2	Emotion or abstraction A2
Where?	Individually named geographical location S3	Kind of place: geographical, architectural G3	Place symbolized A3
When?	Linear time: date or period S4	Cyclic time: season, time of day G4	Emotion, abstraction symbolized by time A4

The Panofsky/Shatford facet matrix has become a widespread model for describing image content. Several studies have used this matrix in classifying image queries or requests. Studies have determined that queries and requests include multiple facets, on average one-and-a-half to two facets per query (Armitage & Enser, 1997; Choi & Rasmussen, 2003; Enser & Sandom, 2001). In these studies approximately 60% of the facets have been specific, while less than 40% were generic and less than 5% were abstract. Facet analyses suggests that especially the specific *who* (Who is portrayed?) and *where* (Which location is portrayed?) facets seem to be pronounced as is the generic *who* (What type of a being is portrayed?) facet.

Images may also be queried based on syntactic properties. Keister (1994) classified image requests into two categories: 1) visual requests which define elements that should be seen in the image and 2) topical requests with no specific visual requirements. In Keister's sample up to a half of the requests was based on visual constructs. Later studies have indicated that purely perceptual and compositional terms only account for a small percentage of all image queries (Jörgensen & Jörgensen, 2005, Hollink et al., 2004).

Nonvisual attributes may serve as access points to both syntactic and semantic image content. They are also crucial in image searches concerning the context of the production of the image, such as a known photographer, format, or identifier (Jörgensen, 1999). Furthermore, image needs may refer to a specific item (Conniss et al., 2000; Jörgensen, 1999). These needs may be satisfied only by a single known image (e.g., Le Baiser de l'Hôtel de Ville, Paris, by Robert Doisneau) while others may be fulfilled with various images (e.g., a couple kissing).

Nearly half of image queries are modified with refiners concerning for example time, location, actions, events or technical attributes (Enser, 1993; Markkula & Sormunen, 2000). In web searching the use of refiners may be even more pronounced, especially with nonunique image needs (Jansen, 2008). Refiners may refer to the whole image or to the objects in it. The refiners may serve to refine general terms (e.g., girl) into a more specific visual (e.g., blonde girl) or abstract requests (e.g., beautiful girl) (Goodrum & Spink, 2001). Syntactic image attributes are sometimes used as refiners in otherwise conceptual queries, adding specifications of for example shooting distance (Markkula & Sormunen, 2000).

Little work has been conducted in comparing and validating image query typologies. Chen (2001) has compared the schema of Enser (1993) with the task types of Fidel (1997) and the image description template by Jörgensen (1998). These frameworks were determined to be applicable in categorizing art image queries but there were issues in agreement between multiple judges. Also, the results differed from the original findings because of user and domain characteristics, which shows the importance of domain-specific research. Based on Chen’s work Jansen (2008) characterized web image queries concluding that the existing classification schemas did not port to the web image searching environment. They were not able to capture attributes associated with web image queries (e.g., URL, genre, cost).

2.3 Previous Work on Journalistic Image Access

Image search strategies of newspaper journalists have been investigated in various case studies. Markkula and Sormunen (1998, 2000) investigated journalists' illustration tasks, queries and requests to the image archive. The goal of image retrieval was to make the illustrated page look attractive, balanced and dynamic. Browsing was the main search strategy, and it helped journalists to develop illustration ideas and to employ dynamic relevance criteria. The journalists relied on simple single word or phrase queries. They had difficulties finding photographs of generic objects or themes and relied on intermediaries in the archive, requesting photographs. Based on the findings, Markkula and Sormunen suggested changes to the indexing procedures and improvements to the retrieval system to better support end-user queries.

In specialized image collections intermediaries play an important role in the description and retrieval processes. Intermediaries must form a model of the end-user's image need, receive feedback on relevance, for example, and guide the search process based on that feedback (Ingwersen, 1992). There exists evidence from text retrieval that intermediaries affect users' interaction with the retrieval process (Spink, Goodrum, & Robins, 1998). Ørnager (1995, 1997) studied the work of intermediaries in newspaper photographic archives. Based on interviews and observations of journalists formulating requests, Ørnager proposed a typology of image searchers. The specific inquirer formulates a narrow request based on a specific image she has in mind. The general inquirer formulates a broad request, wanting to choose without interference from archivists. The story-teller inquirer discusses the topic and is open to suggestions. The story-giver inquirer gives the article to the archivists wanting them to choose. Finally, the fill-in-space inquirer wants a certain size photograph to fill the page. Based on these searcher types, Ørnager (1997) suggests retrieval solutions ranging from conceptual query reformulation to browsing tools for finding similar images.

Preferred access points to images have also been analyzed previously through surveys and analysis of expressed image needs in journalistic contexts. Ørnager (1995) surveyed image indexing and search needs at newspaper photograph archives. She lists the minimum requirements for image indexing in this context: named person (who), background information about the photo (when, where), specific events (what), and moods and emotions shown or expressed, as well as the size of the photo. Armitage and Enser (1997) analyzed image requests finding that most requests from magazine and newspaper publishers were specific, requesting images of named objects, locations or events. Markkula and Sormunen (2000) determined in their case study that most image needs of journalists were related to photos of persons and other named objects. Queries were also often concerned with recent news events while requests for types of objects were common. Neal (2008) administered a survey on image access points to photographers, photojournalists and news librarians, extending the data with follow-up interviews. Using the typology from Burford et al. (2003) to loosely structure the answer alternatives, Neal found that photographer-assigned keywords, basic photographic metadata properties, named objects and events were thought to be the most useful for retrieving news images. Sormunen et al. (1999) studied the similarity judgments made by journalists, in order to evaluate content-based image retrieval solutions. The similarity criteria ranged from visual characteristics of thumbnail images to affective factors interpreted from enlarged photographs. Overall, journalists employed a wide range of syntactic (e.g., color, background) and semantic (e.g., action, atmosphere) similarity criteria when evaluating images.

3 Methods and Experiments

The overall goal of this dissertation is to understand the processes in journalistic image access as well as the effects of contextual factors on the preferred access points to journalistic images. The study of image attributes, categories, queries and searches requires distinct and diverse methods, as discussed in the review of related research. These methods were applied in the separate studies that form this dissertation. The studies build on both quantitative and qualitative data. A literature review on image attributes was conducted, to form the basis of an image attribute typology. Image attributes were also analyzed from empirical data gathered in two description tasks. Image search processes and query types were analyzed based on a newspaper field study and a laboratory experiment in multimodal image retrieval. Image categorization criteria were studied in multiple experiments, leading to a categorization model. In the multimodal image retrieval and categorization studies advances on research methods were implemented. All of the studies were conducted in Finnish and the participants were native Finnish speakers.

3.1 Image Attributes

An experiment in image description with two different description tasks was conducted (I). The image descriptions were analyzed for image attributes. Existing theoretical and empirical frameworks of image attributes (II) were the subject of a literature review and grounded theory analysis aiming for a typology of image attributes for image access.

3.1.1 Image Attribute Elicitation

The experiment sought to find the image attributes prevalent in the description and keywording of journalistic photographs. Possible differences between the attributes elicited in the two tasks were also investigated.

The test material consisted of 40 reportage-type photographs depicting people in various situations, inanimate objects, animals and scenery. The images were gathered from two online image collections by image journalists and photographers. Several criteria were used for the selection: a broad range of color distribution, colorfulness (Hasler & Süsstrunk, 2003) and lightness levels, strong visual elements, various distances to object, and a wide range of topical and emotional content.

A total of 20 participants (eight female) were recruited consisting of students and university employees. Participants were divided into two groups at random and gender balance was maintained. One group performed the free description task and the other the keywording task. The photographs were displayed on a computer screen in random order without time limitations. In the keywording task the participants were asked to write the first five words that came to their minds that best described the photograph that they were seeing. In the free description task they were asked to write a description of the photograph as they would offer when describing its content to another person. The keywording data consisted of five individual words or multiword terms per image. In the free description task, an unconstrained description was requested. The participants mostly wrote complete or near-complete sentences, from which on average eight descriptive terms were extracted.

The terms from both of these tasks were categorized according to Jörgensen's (1998) image attribute framework. The chi square test was used to analyze the differences between the attribute distributions in the two tasks. Standardized residual analysis was used to determine which categories were major contributors to the statistical differences.

3.1.2 Theoretical Review and Typology

A review of image attributes (II) was conducted based on empirical studies and theoretical work (Burford et al., 2003; Eakins et al., 2004; Hollink et al., 2004; Jaimes & Chang, 2000; Jörgensen, 1998; Shatford, 1986). A grounded theory analysis was done by coding and comparing attributes and attribute levels in the frameworks. The end result was a typology of image attributes on several levels.

3.2 Image Search Process

Image search processes were studied in two different contexts: a newspaper field study (III) and a laboratory experiment on multimodal image search (IV). The image searching behavior of journalists and intermediaries was analyzed based on a qualitative case study. The goal was to model the image selection process, including the search process and the relevance criteria, in the context of journalistic work tasks. In the area of multimodal image retrieval, search patterns were analyzed from expert and nonexpert users' interactions with a multimodal image retrieval system. The analysis also aimed at discovering the effects of the task or user type on the search tactics.

3.2.1 Journalistic Image Selection Process

The image searching behavior of the staff at a daily Finnish newspaper was analyzed. The newspapers' editorial system was the main tool in the image workflow, including image retrieval. Images were sought by the journalists and by archivists acting as intermediaries. Images could be queried by free text and by specifying the photographer, shooting/publishing date, source, location, photograph direction, or shooting distance. The digital image archive of the newspaper held approximately 300 000 photographs at the time of the study.

Nine theme interviews with key actors of the image selection process (journalists, image journalists, photographers, editors, graphic artists, and archivists) were conducted. Interview themes included work tasks, image needs, the retrieval process, selection criteria, textual information, editorial system, and collaboration. The image-related work tasks of the interviewees were also observed either before or after the interview. In addition to this general observation, six image selection processes were observed. The qualitative data gathered from the interviews and observations were analyzed in several steps by transcribing, coding and classifying meaningful expressions. These were used to model the image retrieval process.

3.2.2 Multimodal Image Search Tactics

To analyze the search tactics in a combination of concept- and content-based image retrieval a user test with a prototypical multimodal image retrieval system was conducted. The aim was to determine whether and how the multimodal access modes would be combined in queries.

Two groups of users participated in the test. The seven expert users (six female) were on average 34 years old (*SD* = 9). They had on average eight years of experience performing image retrieval tasks in magazines with job descriptions ranging from art director to editor. The seven nonexpert users (all female) were university students in various fields. They were on average 23 years old (*SD* = 2) and none had professional experience in image retrieval.

The first contribution of this study is the developed set of six image search tasks. Simulated multimodal search tasks were used because of the need to control variation. The tasks were designed based on reviewed literature and previous work (III). The order of the tasks was randomized and participants were told to complete the tasks at their own pace. The task instructions were as follows:

1. Find the attached photograph (of a model car).
2. Find out which country has a black, red and white striped flag with a bird in the center.
3. Find an image of a green field with blue sky.
4. Find a high quality nature image of snow-covered trees with white as the main color.
5. Find 2-5 images to illustrate an article on a shopping holiday in Rome.
6. Find an image which portrays hurry and quest for success to illustrate an article on business.

The task instructions included both conceptual and visual requirements. Most of the tasks required one image to be returned. Task 1 asked for a specific item, and task 5 asked for a selection of images. Unfortunately, it was not feasible to include a task for finding an image of a specific person because the collection being used did not include such material. Table 2 maps the tasks to the image search task typology by Jörgensen (1999). In addition to object tasks, a data task (Fidel, 1997) was included. The expert participants evaluated the validity of the simulated search tasks. When asked whether magazine journalists, archivists and photographers perform tasks such as these (1 = completely disagree, 5 = completely agree), the participants overall agreed (*M* = 3.8, *SD* = 1.1). They themselves also typically performed similar tasks and evaluated the task instructions as clear (scale 1-5) (Table 2).

Table 2. Tasks typified and evaluated

Task	Mean clarity	Mean typicality	Task type	Jörgensen (1999)	Fidel (1997)
1. CAR	5.0	3.7	Known item/ data	Specific item	Object
2. FLAG	3.7	3.6		Specific instance	Data
3. FIELD	4.9	4.4	Visual cues	General category	Object
4. TREES	4.9	4.3		General category	Object
5. ROME	5.0	4.3	Conceptual/ Abstract	Specific instance	Object
6. HURRY	4.4	4.4		Abstract/affective	Object

A prototypical multimodal image retrieval system was implemented in this experiment due to the requirement of including various query modes (text, color, sketch, quality, category, similarity queries) and a workspace. This system also enabled transaction logging of all interactions. The system contained 7500 keyword annotated stock photographs organized into a topical category tree. The topic areas of each task were well represented within the collection. To verify baseline usability, users evaluated the system at the end of the experiment. On a scale from 1 to 5 (1 = completely disagree, 5 = completely agree) participants thought that the system was useful (M = 4.3, SD = 0.5), efficient (M = 4.1, SD = 0.5), easy to learn (M = 4.4, SD = 0.6), pleasing (M = 4.0, SD = 0.9) and rather reliable (M = 3.9, SD = 0.9). The participants found the idea of combined, multimodal queries simple to understand (M = 4.2, SD = 0.9) and their construction was easy (M = 4.1, SD = 1.0).

Transaction logging was used to gather data on user interactions with the system. Statistical analysis of the effects of user and task type on query modes and modifications was performed with chi-square tests and standardized residual analysis. Query modifications were sorted based on transitions from one query to another into the following categories: additive (adding query modes), subtractive (reducing query modes), switch (switching query modes, called equal in IV) or repeated (repeating the same query mode(s)). Search patterns were analyzed based on the likelihoods of transitions from one interaction to another, for example moving from a text query to viewing a result image. Different degree Markov models show which interactions are likely to follow one another. Maximal repeating patterns in user interactions were also indentified. A maximal repeating pattern (MRP) is a repeating pattern as long as possible or a substring that occurs independently (Siochi & Enrich, 1991). The log data from experts were concatenated into one string and the data from nonexperts into another string. Markov and MRP analyses were performed on these strings to identify patterns across users.

3.3 Image Queries and Requests

Users' image needs as expressed in requests and queries were analyzed for the image attributes to which they referred, and for their structure. Publication III describes the journalists' image needs in the newspaper context while IV focuses on multimodal image queries in an experimental setting.

3.3.1 Journalistic Image Query and Request Types

In the newspaper field study (III, Section 3.2.1), image needs expressed as queries or requests were analyzed. Content analysis was performed on the requests received by archivists and on the queries made to the digital image collection. Image requests (N = 66) made by the editorial staff were gathered by the archivists. The requests were analyzed for their semantic content and were categorized according to the typologies by Enser (1993), Keister (1994), Fidel (1997), and Jörgensen (1999). The facets present in the requests were classified according to the Panofsky/Shatford matrix (Armitage & Enser, 1997). Image queries (N = 1788) were obtained from search logs in the editorial system. Queries were analyzed only in relation to some of the typologies because they lacked information about search motives and on whether the queried image was a specific item previously known to the searcher. Image queries and requests were compared.

3.3.2 Multimodal Image Queries

An analysis of multimodal image queries (N = 439) was performed based on user interactions logged in the multimodal image search experiment (IV). The query modes (text, color, sketch, quality, category, similarity queries) and their combinations used in the search tasks were analyzed. Possible effects of task type (known item and data, visually cued, conceptual and abstract) and user expertise were also investigated. Statistical analyses were performed with chi-square tests and standardized residual analysis. The participants and tasks are described in Section 3.2.2.

3.4 Image Categorization

Subjective image categorization tests were conducted to elicit categorization criteria. Based on these tests, a framework of magazine image categorization criteria was developed and evaluated. Study one (VI) elicited subjective magazine image categorization criteria from expert and nonexpert participants. The effect of the page context of magazine images on experts' categorization behavior was further investigated (V). Study two (VI) evaluated the framework in a separate categorization task. Connections between classes were analyzed based on their use, and were included in the final categorization model.

3.4.1 Image Categorization Criteria

The participants in the categorization study were divided into three groups. Participants were first divided into nonexperts and experts, based on their knowledge and experience in image categorization. The nonexpert participants (eight female) were university employees or students. Their average age was 24 (SD = 5.2). The expert participants (23 female) were staff members at magazines, newspapers, picture agencies and museum photograph archives. Their average age was 43 (SD = 8.6). The expert participants were further divided into two equal groups: those categorizing images in page context (context group) and those categorizing images without context (no context group). Experts and nonexperts were compared in VI while effects of context were analyzed in V. The dataset for the expert group in VI is the same as the no context group in V.

The 100 test images were selected at random from magazines of different genres (women's, economy, general, lifestyle, travel). Twenty images were taken from each of the five magazines. Two material sets were created. For the context group the images were left into their page context while for the no context group and nonexperts the images were cut out of the page context and any remaining text elements were covered. The experimental procedure consisted of two phases for all groups. In phase one the participants were instructed to sort the photographs into an unrestricted number of piles according to their similarity, such that photographs similar to each other would be in the same pile. The participants were told to decide on their own the basis on which they would evaluate similarity. There was no time limit. The participants were subsequently asked to describe the similarity element in each pile, i.e. to name the piles. In phase two the participants were shown the category names from phase one. They were asked to go through the photographs again one at a time and assign each photograph to suitable categories (one or more).

The category names supplied by the participants were analyzed qualitatively, to construct classes of categorization criteria. The analysis was data-driven but it was informed by theory and previous research on image categorization as reviewed in this dissertation. After iterative analysis, main classes of image categorization criteria were defined, with identifiable subclasses, and were gathered into an image categorization framework. The category names from the experiment were coded according to the classes. A single category (a group created by a participant, for example, *people and work*) could include references to multiple classes (coded People-Person and Theme-Work according to the framework) resulting in a multi-class category. Results are reported on the frequencies of different classes in the category names by the groups. Statistically significant differences are based on chi-square tests and standardized residual analysis.

Multidimensional scaling (MDS) was used to investigate the dimensions utilized in similarity evaluations. For this analysis data from the two no context groups, both expert and nonexpert, were used. Data from phase one were used because it contained the primary similarity judgments reflecting the employed similarity dimensions. The similarity of two photographs was calculated using the percent overlap measure (Rorissa & Hastings, 2004) which reflects the share of participants who placed the two images into the same category. MDS has also previously been used to analyze similarity dimensions in image sorting (Rogowitz et al., 1998; Teeselink et al., 2000). The stress value (Borg & Groenen, 2005) was used to evaluate the fit of the scaling solution to the original data.

Hierarchical cluster analysis was used to analyze whether the page context affected the categorization by experts. Data from phase two were used because it contained all of the available similarity judgments within the image set. Because of the procedure in phase two which allowed for multiple categorizations of the same image, a modified percent overlap measure was used to calculate image similarity. The overlap measure was normalized by the overall number of categorizations made for the images. Jaccard's coefficient was used to measure whether categorization differed between the groups. This measure has been used by Lohse, Biolsi, Walker, and Rueter (1994) as well as by Rorissa and Hastings (2004) to test consistency in image categorization tasks.

3.4.2 Image Categorization Framework Evaluation

The image categorization framework was evaluated in a task-oriented manner in Study two (VI). The framework was embedded in an image archiving application which expert participants used to categorize a set of stock images. Both the categorization behavior and the subjective evaluations of the framework were analyzed. The goal was to demonstrate the usefulness of the framework in a categorization task and to gather information on its application. Based on these results, the framework was extended into a categorization model.

Differences were found between experts and nonexperts in the types of categorization criteria employed (Study one, VI). For Study two only domain experts were recruited. The 24 participants (18 female) were on average 37 years old (*SD* = 8.5). The participants had several years of experience (*M* = 9.5, *SD* = 8.8) in image retrieval, categorization or indexing. Their job descriptions in magazines ranged from archivists to journalists, graphic designers and photographers.

The 20 test images were drawn at random from a laboratory test pool of 2500 stock images. Participants were shown three practice images followed by the test photographs in random order, together with the categorization framework. The participants were given a simulated work task (Borlund, 2003) asking them to give descriptive categories to the photographs such that other employees could find them by browsing the archive. The photographs were to be categorized according to their essential content. Within the interface the participant saw one image at a time and was to select appropriate categories from a check-off list. Participants were free to browse back and forth between the images, to correct and complete their categorizations. Selections could be further specified in a text box that opened when the category was checked. The categories corresponded to the subclasses discovered in V and VI with these additions based on the larger pool of studies (Laine-Hernandez & Westman, 2008; 2010): subclass Nature Object (main class Object), Science (Theme), Repetition (Visual), Impression (Visual), and Perspective (Photography).

After categorizing each image, participants completed a survey that included questions about the perceived difficulty of the categorization task (hard-easy; slow-fast) and the fit of the categories to the image (poor-good) on 5-point Likert scales. Participants also answered a post-test questionnaire and were interviewed about their experience using the framework. They were asked whether each main class had been useful in categorizing the images and whether the classes were well-defined, on a 5-point scale (1 = completely disagree, 5 = completely agree).

Rolling's (1981) interindexer measure was used to depict participant agreement in categorizations at the subclass level. The measure was averaged across all participant pairs for each photograph, to obtain an average measure of consistency. Dice's measure (van Rijsbergen, 1979) was used to evaluate the coherence of the classes as reflected by the placement of the same images into the same subclass by different users. The measure was calculated for the 18 subclasses which were used by at least five participants.

3.4.3 Final Image Categorization Model

Connections between the classes in the framework were analyzed based on the participants' strategies for combining them in category names or selections and were included in the final model. The connections were analyzed based on co-occurrences of class instances in multiclass category names in Study one, overlap in category use in phase two in Study one and combined category selections in Study two. These analyses shed light on the connections between the classes in users' assignment of them and were used to transform the categorization framework into a model. The strongest symmetrical connections normalized by the overall use of the classes are presented.

4 Results and Discussion

The key results of the studies are reported based on contributions to the dissertation in the areas of image attributes, image search processes, image queries and requests, and image categorization. The results are discussed in relation to previous work, including studies reviewed in Section 2.

4.1 Image Attributes

Attributes of images were elicited in an experimental setting (I) and were reviewed based on literature (II). A typology of image attributes was created, and an established classification scheme (Jörgensen, 1998) was utilized in the experiment to analyze the descriptions obtained. Nonvisual, syntactic and semantic attributes were distinguished in the typology based on previous work. The prevalent description level in both description tasks in the experiment was the generic semantic level. Differences were found based on the description task issued (keywording vs. free description).

4.1.1 Image Attributes in Description Tasks

Image attributes in the free image description and keywording tasks were classified according to classes from the typology of Jörgensen (1998). Images were described with terms related to the object(s) visible in the image, the story conveyed in the image, and various terms related to the object(s), scene and people portrayed (Table 3). Most of the terms used in the description and keywording of the photographs were semantic and generic. This is consistent with results from earlier image description studies (Fidel, 1997; Greisdorf & O'Connor, 2002a; Jörgensen, 1998).

The distribution of the attributes for the keywording and free description groups differed significantly on the main class level ($\chi^2(10) = 590.36$, $p < .0001$). The difference was contributed to the most by the different use of location attributes, content and story attributes, abstract attributes as well as terms related to viewer response. All attributes other than external relation showed significant differences. Keywording resulted in more terms depicting story, setting and theme than free description, which led participants to enumerate individual objects and describe their locations. Free description also resulted in narrative-type descriptions which could include conjectures and estimates. Attributes referring to theme, event and setting, atmosphere and emotion were used more in the keywording task, which reflects the need to summarize interpretations. The use of numbers, colors and other such descriptions was less common in the keywording task, which was also likely a result of the need to express the content concisely. Visual elements such as motion were mentioned more as keywords. This seems to be the result of summarizing the photograph rather than recounting its content part by part.

Table 3. Percentage distribution of attributes in the tasks

Class	Attribute	Keywording		Free description	
Objects	Object	22.8	26.3	19.5	29.1
	Text	0.0		0.3	
	Body part	2.0		6.0	
	Clothing	1.5		3.4	
People	People	4.1	4.1	7.0	7.0
Color	Color	2.5	3.0	4.4	6.2
	Color value	0.5		1.8	
Visual elements	Composition	0.4	6.4	0.6	4.0
	Focal point	0.4		0.2	
	Motion	1.9		0.4	
	Orientation	0.6		0.3	
	Perspective	1.7		1.3	
	Shape	0.6		0.6	
	Texture	0.1		0.1	
	Visual component	0.9		0.4	
Location	General	0.2	0.3	4.4	10.2
	Specific	0.1		5.8	
Description	Description	7.1	7.4	8.5	12.0
	Number	0.4		3.4	
People-related attributes	Relationship	1.3	12.1	0.5	8.7
	Social status	8.6		7.8	
	Emotion	2.2		0.4	
Abstract	Abstract	3.9	10.8	1.0	1.7
	Atmosphere	1.0		0.2	
	State	0.8		0.5	
	Symbolic aspect	0.0		0.0	
	Theme	5.2		0.0	
Content/ story	Activity	8.9	28.2	9.6	17.4
	Category	0.0		0.2	
	Event	6.7		1.8	
	Setting	10.5		5.0	
	Time aspect	2.1		0.8	
External relation	Comparison	0.0	0.4	0.1	0.3
	Similarity	0.1		0.1	
	Reference	0.3		0.1	
Viewer response	Personal reaction	0.7	0.9	0.5	3.6
	Conjecture	0.2		2.2	
	Drawing	0.0		0.0	
	Uncertainty	0.0		0.8	
Total		100	100	100	100

The difference between free description and keywording suggests that limitations imposed by description tasks (e.g., separate terms, limited number of terms) may truncate natural image descriptions. Jörgensen (1998) also discovered effects of the description task on the attributes utilized. Terms related to viewer response (e.g., personal reactions and assumptions, expressions of uncertainty) were more common in free descriptions than as keywords. Lee and Neal (2010) also found a similar effect related to these types of descriptors in describing and indexing a set of images on the web.

The image attributes from Jörgensen (1998) were determined to be applicable for the analysis of these descriptions of journalistic images. In addition, descriptive terms related to animals and weather phenomena were found, and evaluations of visual quality were discovered. Settings and events were commonly described both on the generic and specific levels. No art historical attributes were found in the data. Content and story attributes were more frequent in these results than in those of Jörgensen (1998), as participants more commonly described activities within the photographs. Also abstract concepts and people-related attributes were more common than previously found, whereas color and location attributes were used less. This difference could arise from differences in the format and genre of the images used, as Jörgensen employed illustrations. Greisdorf and O'Connor (2002a) reported that viewers utilized content-based descriptors when imposed on them, but not when asked to come up with their own descriptors. However, the results show that a fifth of the terms in the free description task were syntactic in nature. Syntactic attributes alone are not considered sufficient for indexing photographs. Within a single semantic content category, however, syntactic attributes could aid browsing.

4.1.2 Typology of Image Attributes

In I, a preliminary classification of access points to images was drafted based on literature (Hollink et al., 2004; Jaimes & Chang, 2000; Jörgensen, 1998; Shatford, 1986). This classification was further extended (Burford et al., 2003; Eakins et al., 2004) in II to form the image attribute typology (Table 4). The classification brings together the types of attributes identified in images across the reviewed frameworks. The far left column provides a summarizing classification of image attribute classes. Images may be described, indexed and queried with attributes on three main levels by using nonvisual, syntactic or semantic image information. As one goes down the table, each listed syntactic and semantic attribute type represents a higher level of abstraction than its predecessor.

Nonvisual image information refers to information not present in the image itself but rather associated with its production and presentation. Nonvisual information may contain bibliographical attributes (e.g., creator, date), physical attributes (e.g., type, technique) or contextual attributes (e.g., caption, reference). These attributes often take the form of metadata.

Syntactic image information (image syntax) is information present in the visual characteristics of images. Syntactic attributes may address three levels of image syntax. Global distribution refers to the image-wide distribution of low-level visual content such as color or sharpness. Local structure addresses nonrepresentational image components, such as shapes. Image composition refers to the spatial layout of the components.

Semantic image information (image semantics) refers to conceptual image content. Semantic attributes may be generic, specific or abstract. Generic semantic attributes describe types of objects or scenes, while specific attributes refer to identified and named objects or scenes. Abstract attributes refer to what the image represents, such as its symbolic aspects, interpreted by the viewer or her emotions elicited. Semantic attributes on any level may refer to various issues, including people, objects and settings.

Table 4. Typology of image attributes. Adapted from II. Copyright 2009 John Wiley & Sons, Ltd

Attribute type		Panofsky/ Shatford (1986)	Jaimes & Chang (2000)	Hollink et al. (2004)	Burford et al. (2003); Eakins et al. (2004)	Jörgensen (1998)
Nonvisual	Bibliographical		Bio-graphical information (Nonvisual)	Date, material, style/period, culture, ID number, title, creator, rights (Nonvisual) Type/ technique (Perceptual)	Metadata	Artist, format, medium, time reference, style, type, technique, representation (Art historical information)
	Physical		Physical attributes (Nonvisual) Type/ technique			
	Contextual		Associated information (Nonvisual)	Relation (Nonvisual)		Comparison, similarity, reference (External relation)
Syntactic	Global		Global distribution	Color, texture (Perceptual) Shape (Perceptual)	Perceptual primitives Geometric primitives	Color, color value, texture (Visual elements) Shape (Visual elements)
	Local		Local structure			
	Compositional		Global composit-ion	Composition, position, spatial relation of elements (Perceptual)	Visual relationships, Visual extension	Composition, motion, orientation, perspective, focal point, visual component (Visual elements)
						General/specific location (Location)
Semantic	Generic	Pre-iconography generic "of"	Generic object	General conceptual level: object, scene, event, place, time	Semantic units	Object, body part, clothing, text (Objects) People (People) Activity, time aspect, event, setting, category (Content/story) Number, description (Description)
			Generic scene			
	Specific	Iconography specific "of"	Specific object	Specific conceptual level: object, scene, event, place, time		
			Specific scene			
	Abstract	Iconology "about"	Abstract object Abstract scene	Abstract conceptual level: object, scene, event, place, time	Contextual, cultural and technical abstraction	Abstract, state, symbolic aspect, theme (Abstract concepts) Relationship, social status (People-related attributes)
					Emotional abstraction	Emotion (People-related attributes) Atmosphere (Abstract concepts)

The number and types of attributes required to convey the content of an image are still under debate. The meaning of an image may be seen to emerge from a user's interaction with the image collection, making the appropriate description levels dependent on both the collection and user needs (Santini, Gupta, & Jain, 2001). In any case, the image attributes considered useful in a specific retrieval context should be offered as access points to searchers. This may be achieved by intellectual indexing, the analysis of image features or a combination thereof, depending on the nature of the access points.

4.2 Image Search Process

The image selection process modeled at the newspaper (III) was found to be highly contextual. The search goals and strategies depended on the work task which also gave rise to the relevance criteria. The image need was conceptualized into a mental model of the image sought, against which the retrieved and browsed images were compared. In multimodal image retrieval (IV) the search task type had an effect on the search interactions. Query modification patterns differed for nonexpert and expert searchers.

4.2.1 Journalistic Image Selection Process

The image selection process observed during the newspaper field study (III) is modeled in Figure 1. This model of a journalistic image selection process confirms and extends the model of journalistic illustration processes by Markkula and Sormunen (2000). It synthesizes image seeking activities in the selection process by providing a joint model of the search for archived images and the acquisition of new photographs through shoots. It also includes different browsing approaches, which combine visual and textual information.

The goal of the image selection was to select the best image in relation to the illustration task. Norms and practices could impede the selection of the absolute best image. The utility of the image was emphasized especially when the layout was not known at the time of selection. Images were then selected to suit the envisioned layouts. Often an image that would work in most publication scenarios was selected as opposed to the best image for each scenario. The goal of finding the most useful image in the specific illustration task reflects relevance on the situational level (Saracevic, 1997). Depending on the role of the person selecting the images, the images were thought to have several functions: acting as proof of a news event, catching the readers' attention, filling up pages, conveying information or adding value by bringing a new aesthetic or informational dimension to the article. Image retrieval was also performed to complement factual information retrieval.

Upon receiving an image request or an illustration task, the searcher formed a mental model of the image sought based on the task. The factors included the article type, the paper section and the position in the layout. The model was also influenced by previous information on the topic and expectations regarding available images. If suitable images were not available in the archive, an internal image order was made to photographers. These image orders typically combined conceptual and visual criteria for the images to be taken. In mediated search, the archivists' model of the images sought was influenced by factors related to the journalist formulating the request. Image needs were specified after seeing initial retrieval sets as the perception of the illustration task changed, and new illustration possibilities were discovered.

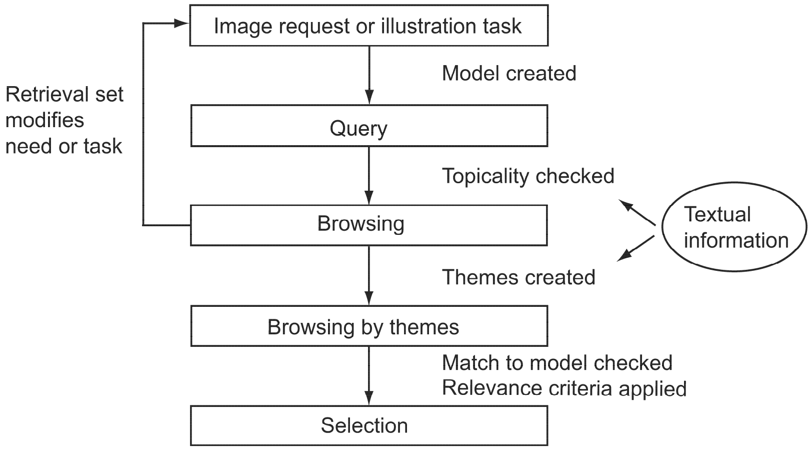

Figure 1. Model of the image selection process. Adapted from II. Copyright 2006 ACM

Simple text queries were used to begin searches. Image needs were often fuzzy and sometimes could not be explicated beyond naming critical objects that should appear in the image. When evaluating retrieved images, the searcher first ensured that the image was related to the search topic based on the visual content, the attached textual information or the viewing context. Browsing was the main search strategy after the initial query and was especially important in abstract searches and collaborative retrieval. Both the retrieved and the recently captured images were compared against the mental model of the image sought. During browsing, searchers seemed to mentally categorize the images into thematic categories, as exemplified by the following quote:

> *I have three themes here, which I am considering depending on the layout, on how they construct the article. I will select example photographs of each.*

This categorization may also be made explicit, for example when browsing image categories by an image agency. Themes were created by all actors observed, including the photographer classifying her photos into themes using mostly compositional criteria. While browsing images within a theme, searchers focused on features that separated the images from each other or that especially caught their attention. Images were sometimes selected from a large set by a gradual process of paired comparisons. These strategies seemed to help structure the task and to create a limited number of dimensions where the search could be focused.

Several types of criteria were used in the relevance assessments made to select an image for publication. Contextual factors (e.g., the nature of the article, publishing section, available space and page layout) formed a selection frame for suitable images. Topicality was a necessary, but insufficient, criterion for relevance and it was used as a starting point. Compositional and informational criteria followed in later stages of the selection process. The final selection criteria were dynamic, activated by comparisons of retrieved images and based on the characteristics and differences between them, as is apparent from this quote:

This one is more static, this one has depth, they are both sharp.

The final selection criteria could also be preferential or reactive; selections were based on impressions of images being, for example, “more interesting", “funny", “different", or “most dramatic". Several implicit criteria, similar to those reported by Conniss et al. (2000), were employed in the selection process: recency, accessibility, cost, and close-ups for photographs of people. Constraints such as previous publication and presence of other images on the page or in the article also influenced the selections. The role of textual descriptions in relevance judgments was not always emphasized in interviews, however, during observation it was noted that searchers often used associated textual information to verify topicality. This type of relevance assessment has also been previously reported (Frost, 2001; Markkula & Sormunen, 2000). Searchers tended to alternate between viewing the textual description and the image during the selection process. Text was especially important when the search topic was previously unknown.

Several types of collaboration were observed in the image selection. Some journalists always gave their illustration tasks to archivists or searched with them, while others searched by themselves or requested help based on the situation. Personal preferences, their own skills and experience, the size of the image set and the schedule affected the journalists' choice on when to request help. Knowledge about searchers and searcher types helped archivists understand their image needs. Multiple image searcher types could be identified, following the general lines of Ørnager's (1997) typology. Some journalists asked for a set of images from which to select, while others preferred that the archivists choose the image. Some wanted a specific photograph to be retrieved. Some journalists communicated image needs briefly, while others provided rich information, even the whole article to be illustrated. Collaboration was also observed between archivists on difficult searches. Image journalists and photographers collaborated on image selection tasks. Photographs were named thematically or were based on salient objects (e.g., circus photographs, motorcycle photographs) to facilitate conversations.

Images were grouped thematically based on topical content, compositional features and contextual factors. These groupings seemed to aid in structuring the image selection task and discussing the image needs in collaborative image retrieval. Browsed image sets seemed to modify the image need or the perceived illustration task and subsequent queries. Garber and Grunes (1992) noted that in art directors' image searches, previously found image sets affected both the so-called image and artistic concepts, i.e., the image need and the work task. Greisdorf and O'Connor (2002a) have also described this phenomenon, casting image retrieval as a process of reconceptualizations in which the concepts in the initial query and the concepts raised by viewing retrieved images merge.

These observations on browsing lend further support to the idea that content-based image analysis could be used in to aid retrieval and to organize retrieval sets for browsing. Support for browsing appears to be crucial in image retrieval interfaces for journalists.

4.2.2 Multimodal Image Search Tactics

Most interactions logged in multimodal image searches corresponded to queries (31%). The remainder consisted of viewing images (23%), resetting query conditions (16%), saving images to the workspace (13%), removing images from the workspace (10%), comparing images in the workspace (4%) and clearing results (4%).

These interaction shares varied by task type ($\chi^2(12) = 120.87$, $p < .001$) (Table 5). The statistically significant difference was brought on by differences in the shares of all the interactions. The known item and data tasks led to most reset query conditions and cleared results, probably due to the specific requirements of the tasks. In visually cued tasks, viewing, saving and comparisons in the workspace were pronounced. Browsing has been determined to be a key strategy for journalists' generic image needs (Markkula &Sormunen, 2000). Current results also highlight the role of image comparisons for generic tasks with specific visual criteria. The conceptual and abstract tasks led to heavy querying and viewing. This reflects the use of a combination of textual and visual access modes, as browsing was used together with query reformulation.

Most first queries (85%) were further modified. The modification patterns suggest that switching query modes was less frequent than adding or subtracting modes. Different types of tasks led to different types of transitions ($\chi^2(8) = 35.71$, $p < .0001$)(Table 5). The statistical difference results mainly from the varying amounts of repetitions of the same query mode. Conceptual and abstract tasks had a large number of repeated queries because users changed query terms and category selections. The visually cued tasks led to most switches as users experimented with sketch, color and similarity queries.

The participant groups also made different types of transitions ($\chi^2(3) = 15.05$, $p < .01$). Experts were more likely to edit the content of their queries (e.g., replacing terms, categories, color values) and resubmit the same type of query (49%) than nonexperts. Nonexperts added (25%) or subtracted (24%) query modes quite often.

Table 5. Common interactions and transitions by task type

Task type	Most common interactions	Most common transitions
Known item/data 1&2	Queries (35%), resetting (28%), viewing images (15%)	Repeating query modes (29%), Subtracting query modes (28%)
Visual cues 3&4	Viewing images (24%), querying (21%), saving images (18%)	Switching query modes (35%), Adding query modes (27%)
Conceptual/ abstract 5&6	Querying (35%), viewing images (26%), resetting conditions (14%)	Repeating query mode (60%), Adding query modes (17%)

Transition analysis enabled recognition of the most probable repeated state transitions (Table 6). Transitions by experts exhibited an iterating or broadening query strategy while nonexperts narrowed their searches by adding more conditions. Most transition patterns were related to issuing a query and viewing result images or issuing another query of the same type. The maximal pattern analysis found common search patterns of different lengths. The length of the average search pattern was 2.6 transitions for experts and 2.5 for nonexperts. The most frequent patterns of all lengths for both groups dealt with inspecting and saving images or querying by text and category. Patterns of inspecting one image, issuing a query and browsing results were also common.

Table 6. Most probable transitions of order N by user group

N	Expert		Nonexpert	
	Move	p	Move	p
1	Browse previous image in result set → Browse previous	83%	Text+color+quality + category query → View one image	83%
	Text+sketch query → Reset sketch	67%	Color+category query → Save image	80%
	Visual similarity query → View one image	60%	Textual similarity query → View one image	67%
2	Browse previous image → browse previous → Browse previous	80%	Reset all query conditions → Sketch query → Text+sketch query	75%
	Text+category query → Text+category query → Text+category query	68%	Text query → Text+category query → Text+category query	75%
	Text+category query → Reset category → Text query	67%	View one image → Textual similarity query → View one image	69%

The multimodal search pattern analysis did not reveal distinct groups of patterns, possibly due to the granularity of the coding scheme. The vast majority of queries were further modified and some search paths were quite long. Expert users issued on average one query more per retrieved image (M = 4.5, SD = 2.9) than nonexperts (M = 3.5, SD = 2.7)($t(82)$ = 1.71, p = .09). Overall, users issued more queries than the typical two to three queries per session found in both web logs (Tjondronegoro et al., 2009) and experimental settings (Goodrum et al., 2003). This result may be symptomatic of the users' lack of understanding of query formulation and strategies to further searches. This issue will to be heightened as more access modes become available.

4.3 Image Queries and Requests

Image queries were analyzed in two contexts: the field study at a newspaper (III) and an experimental setting in multimodal image retrieval (IV). Textual queries were short and focused on specific persons as well as generic objects and topics. Contextual facets of image needs were identified from the requests. Both expert and nonexpert users were able and willing to construct multimodal image queries, drawing sketches, specifying colors, and providing quality criteria as well as entering search terms and category selections. Task type and user expertise were found to affect the queries constructed.

4.3.1 Journalistic Image Query and Request Types

Image requests in the newspaper were determined to be mostly unique (56%) and nonrefined (61%) (as defined by Enser, 1993). They concerned specific items (11%), specific instances (50%), general topics (33%) and abstract concepts (6%) (Jörgensen, 1999). Images were to be used mainly as objects (94%); (Fidel, 1997). The high share (83%) of visual requests (Keister, 1994) is due to numerous requests for images of specific people and types of animals, coded as visual requests.

Image queries included an average of 1.48 query terms, thereby supporting the notion of Markkula and Sormunen (2000) that journalists formulate simple queries. The majority of queries were specific, most often for named persons (40%), organizations (9%) or locations (8%). Generic queries for objects accounted for 17% and generic queries for concepts accounted for 22% of the sample. Abstract and affective queries were less common (5%).

Requests included an average of 1.45 facets (Shatford, 1986) out of which 51% were specific, 45% were generic and 4% were abstract. Of these requests, 86% included a *who* facet, 20% included a *what* facet, 26% included a *where* facet, and 14% included a *when* facet. Queries included an average of 1.07 facets, of which 58% were specific, 37% were generic and 5% were abstract. Of the queries, 77% included a *who* facet, 12% included a *what* facet, 17% included a *where* facet, and 1% included a *when* facet.

There was a difference in the distributions of facets in the requests and queries ($\chi^2(7) = 34.91$, $p < .001$), which was calculated based on the facets which occurred in both (Figure 2). The difference mainly resulted from differences in facets S1 (specific *who*), G1 (generic *who*) and G4 (generic *when*), as names were prevalent in queries while object types and expressions of time were more common in requests.

Also in past studies (Armitage & Enser, 1997; Enser & Sandom, 2001; Choi & Rasmussen, 2003) the facets S1 (specific *who*), G1 (generic *who*), G2 (generic *what*) as well as S3 (specific *where*) were pronounced. These were also the key access points surveyed by Ørnager (1995, 1997). Thus the specific names and types of people and objects serve as important access points to images. Also important are the locations the images depict or were captured in It is noteworthy that generics take precedence over specifics only for events and actions (G2 vs. S2). Based on these studies, abstract facets are rather uncommon in image requests and queries although abstract criteria may play a role at a later stage in the search process.

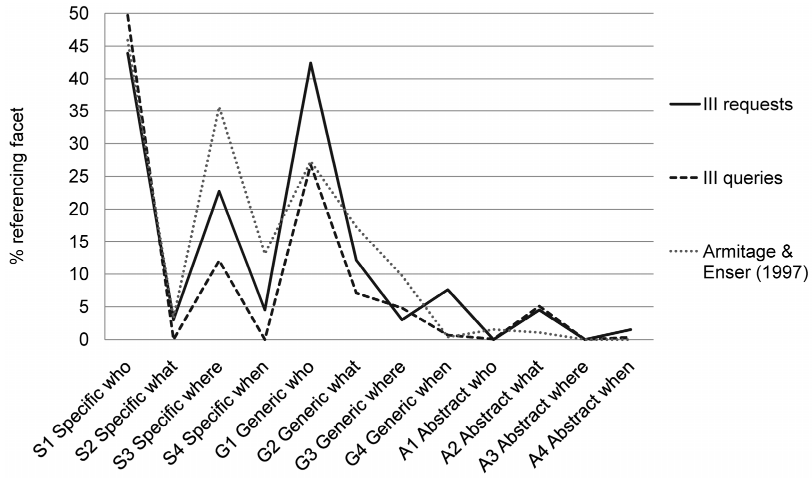

Figure 2. Facets in requests and queries in III and previous work

A comparison of request and query content (Table 7) also resulted in a significant difference ($\chi^2(5) = 26.63$, $p < .0001$), mostly resulting from differences in object type, action or event and thematic requests and queries. Thematic image needs appeared to lead to queries rather than to requests, differing from the results obtained by Markkula and Sormunen (2000). Generic object types (e.g., car) were an important query topic, whereas Markkula and Sormunen only observed them in requests. These differences are probably based on differences in the indexing practices and their development within newspapers in the last decade.

Table 7. Percentage distribution of types of requests and queries

Request or query type	Requests	Queries
Known photo	10.6	-
Named object	37.9	49.1
Named place	10.6	6.9
Object type	31.8	18.3
Place type	0.0	1.3
Action/event	6.1	2.8
Theme	3.0	21.5
Total	100	100

The requests and queries were simple: they included one or two distinct facets and most went unrefined. The low proportion of refined requests compared to previous studies (Enser, 1993; Markkula & Sormunen, 2000) may be caused by some refiners being implicit and dynamically employed during the search rather than featured in the requests. The result may be partly due to changing workflows in which, for example, color images have become the norm. Nearly a tenth of the image requests either explicitly or implicitly asked for all the available material on the topic. This type of request has not received attention in previous studies or typologies.

Upon coding the requests it became evident that not all of the terms could find classification in the facet matrix. These terms referred to the production or publication context of the images. A set of contextual facets for image requests (Table 8) is therefore suggested, following the logic of the facet matrix. A reanalysis of the requests revealed that 18% of facets could be classified as contextual and that over a quarter of requests included a contextual facet.

Table 8. Contextual facets and their percentages in requests

Facet	Description of facet	%
C1 (who)	Photographer	3.0
C2 (what)	Type of image	12.1
C3 (where)	Image source or intended use	13.6
C4 (when)	Publishing time	3.0

It appears that analysis frameworks should be combined or extended to characterize image requests and queries in a journalistic context. They should include the following aspects: semantic level (specific/generic/abstract) of the request, whether the request is visual or thematic and whether it concerns image content or context. The use of images as data for informing the creation of for example graphics should also be indicated.

4.3.2 Multimodal Image Queries

When offered multimodal image access, users were able to formulate visual image queries and combine visual and textual search criteria. Especially expert participants were also willing to edit combined queries and resubmit them. Users combined up to four conditions by using, for example, text, quality, color and category query modes. In this example query the user would input free text, specify relative quality criteria, indicate a color value and select a predefined category of images to search.

Multimodal queries occurred most often in visually cued and known item or data search tasks, while conceptual and abstract tasks led mostly to a single query mode being used. Text was involved in roughly 80% and category selection in 50% of the queries. Queries were short: single text queries included an average of 1.3 terms, and text conditions in combined queries averaged 1.2 terms. Purely content-based searches accounted for only 5% of the queries. However, color mode was combined in 20% of the queries.

The task type had an effect on the types of queries constructed ($\chi^2(8) = 233.75$, $p < .0001$) (Figure 3). Known images and images required for their data were sought with a variety of query modes: text, color, sketches, quality and category. Visually cued tasks produced queries that included color, sketches and quality estimations. Conceptual and abstract tasks led to predominantly textual searches, supported by category selections. Conceptual and abstract tasks led to the longest task times and the most queries. Also, viewing and comparing images was pronounced in the conceptual and abstract tasks. This set of observations is in accordance with Hung's (2005) finding that complex tasks lead to more browsing. The sketch tool was used mostly by nonexpert searchers and more in visually cued tasks. This usage concurs with the conclusions of McDonald and Tait (2003) who suggest sketch tools for retrieving simple images such as landscapes.

There was also an effect of the participant group on the query modes used ($\chi^2(4) = 16.30$, $p < .01$) (Figure 3). Experts preferred text queries combined with color and category selections while nonexperts commonly used a wider variety of combinations, employing, for example, more sketch and similarity queries. The results demonstrate the willingness and ability of searchers to take advantage of visual query modes and construct rich queries. However, as they gain experience in multimodal image searching, search patterns and perceived utility of query modes may change.

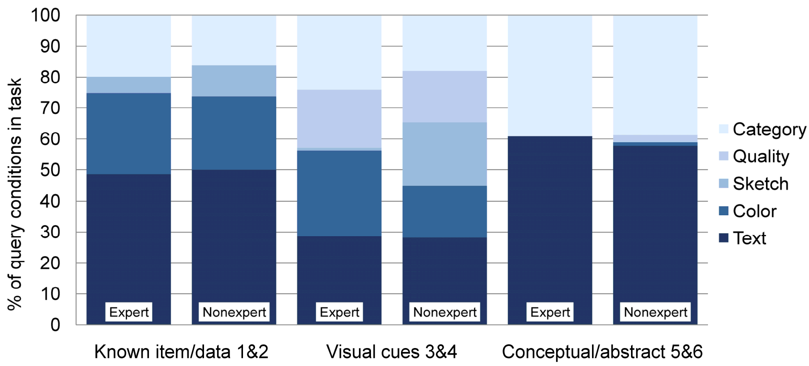

Figure 3. Use of main query modes by task type and user group

4.4 Image Categorization

The categorization of journalistic images was studied in multiple experiments. Ten main classes of categorization criteria were identified, including functional aspects of images, their main semantic content and various types of descriptors (Study one, VI). Differences were found in the application of these criteria between expert and nonexpert users. Publication context was shown to affect category naming (V). A categorization framework was built based on the classes and was evaluated in a study that indicated that, for experts, the distribution of the categorization criteria does not depend on the categorization task or image set (Study two, VI).

4.4.1 Types of Categories

The number of categories constructed varied between 4 and 35. Nonexpert participants created more categories (M = 17, SD = 7) than did experts in either the context (M = 12, SD = 3) or no context group (M = 12, SD = 4). Overall, category names referred to the function of the images, their main visual content (e.g., people, objects and scenes recognized), conceptual content (e.g., theme interpreted), and various types of descriptors (story aspects, affective qualities, general descriptions, photography terminology and visual features).

Table 9 shows the distribution of classes in the category names by participant group. Half of all class instances referred to either People or Theme which were thus the prevalent facets on which magazine images were categorized.

Nonexperts formed categories based on objects in the images (e.g., cars) or that described the scene (e.g., street photographs, European cities). Photographs of people were categorized based on gender, social status (e.g., politicians, public figures), and whether or not the people were posing. Context (e.g., work environment, in the city) was also used as a categorization criterion. Nonexperts also created categories based on emotional impact (e.g., neutral images of people, casual atmosphere).

Experts categorizing images without access to page context typically formed thematic categories such as culture, travel, fashion, work, and transportation. Many categories reflected the photographs' purpose of use (e.g., product photographs, reportage photographs, advertising images). Photographs of people were further categorized based on the number of people (e.g., group shots), social status (e.g., youth/students), relationships (e.g., partnership), what the person represented (e.g., ordinary people) and context (e.g., people at work).

Categories by experts in the context group were predominantly thematic. Several participants formed a category that consisted simply of photographs of people while some distinguished between common people, celebrities and those who represented an organization. Various categories described the function of the photograph (e.g., symbol images, illustrations). The time aspect of the photography was present in some category names (e.g., history).

The main class use differed for the context and no context groups ($\chi^2(9) = 34.06$, $p < .001$), mostly contributed to by the disparity in the use of the classes Theme and Object. Thematic categorization criteria were twice as common in the context group as in the no context group. Without context, four times as many categories as with context were based on the objects depicted. It appears that available page context leads to conceptual, thematic category naming rather than visually oriented, object-based categorization.

There was also a difference between experts and nonexperts ($\chi^2(9) = 54.88$, $p < .001$), which was caused mainly by differences in the use of main classes Function, Affective and Story. Experts created six times as many categories referring to Function as nonexperts. The nonexperts employed the Story and Affective classes roughly three times as often as experts. Nonexpert participants also used the descriptive classes (Description, Visual, Photography) more.

The finding that nonexperts categorized images on the level of story more than experts is interesting since free descriptions of image content often include a story connected with the image (I; Jörgensen, 1998). These results suggest that when categorizing images, experts in the journalistic domain are prone to summarizing story aspects (e.g., time, activity) into thematic descriptions.

Table 9. Percentage distribution of classes in category names

Main/subclass		Context		No context		Nonexpert	
Function	Product photos	1.2	12.3	4.0	19.8	0.7	3.2
	Illustration	3.1		1.7		0.0	
	Symbol photos	3.1		1.7		0.0	
	Reportage	0.6		4.0		0.0	
	Portraits	0.6		2.3		0.7	
	News photos	0.6		1.7		0.2	
	Advertisement	0.6		1.1		0.2	
	Travel photos	0.0		0.0		0.5	
	Misc. function	2.5		3.4		1.2	
People	Person	11.1	19.8	13.6	23.7	8.8	25.2
	Social status	5.6		4.5		3.9	
	Gender	1.2		1.7		4.4	
	Groups	0.6		1.7		1.5	
	Relationship	0.6		1.1		0.7	
	Posing	0.6		0.6		3.2	
	Age	0.0		0.6		1.2	
	Expression	0.0		0.0		0.7	
	Eye contact	0.0		0.0		0.7	
Object	Nonliving	1.2	2.5	4.5	10.7	5.4	10.7
	Buildings	0.6		4.0		3.4	
	Vehicles	0.6		1.7		3.4	
	Animals	0.0		0.6		0.0	
Scene	Interiors	0.6	3.1	2.3	6.8	1.0	10.3
	Landscape	0.0		2.3		2.0	
	Nature	0.0		0.6		0.5	
	Cityscape	0.0		0.0		0.5	
	Misc. scene	2.5		1.7		6.4	
Theme	Food&drink	6.2	50.6	2.8	24.9	4.6	22.2
	Travel	4.9		3.4		1.0	
	Work	4.3		3.4		2.2	
	Transportation	5.6		1.7		0.5	
	Fashion	4.3		2.3		1.7	
	Cinema	3.7		2.8		1.7	
	Sports	3.1		2.8		2.0	
	Culture	3.7		0.6		0.5	
	Art	3.1		0.6		3.2	
	Architecture	2.5		0.6		1.0	
	Technology	2.5		0.0		0.7	
	Hobbies&leisure	1.2		0.0		0.2	
	Home&family	0.6		0.6		0.5	
	Economy	0.6		0.6		0.0	
	Industry	0.6		0.0		0.0	
	Music	0.6		0.0		0.0	
	Politics	0.6		0.0		0.7	
	Religion	0.6		0.0		0.2	
	Misc. theme	1.9		2.8		1.5	
Story	Event	2.5	5.6	2.3	2.8	2.4	7.6
	Time	2.5		0.6		2.4	
	Activity	0.6		0.0		2.7	
Affective	Emotion	0.6	0.6	0.6	1.1	2.9	4.2
	Mood	0.0		0.6		1.2	
Description	Property	1.9	2.5	4.0	5.1	3.7	7.6
	Number	0.6		1.1		3.9	
Visual	Color	0.0	0.0	0.0	0.6	1.2	2.0
	Composition	0.0		0.6		0.0	
	Motion	0.0		0.0		0.5	
	Shape	0.0		0.0		0.2	
Photo-graphy	Distance	2.5	3.1	2.8	4.5	1.5	5.6
	Black&white	0.6		0.6		1.7	
	Style	0.0		1.1		0.7	
	Image size	0.0		0.0		1.2	
	Cropping	0.0		0.0		0.5	
Total		100	100	100	100	100	100

Possible functions of the images were mentioned in nearly 20% of the categories created by the experts, and were also employed by the nonexperts to a lesser degree. The functions included illustrative (e.g., illustrations about places), informative (e.g., documentary images) and persuasive functions (e.g., product images), consistent with literature (Conniss et al., 2000; Pettersson, 1998). This type of functional categorization is based on the idea that specific types of images function similarly when used, rather than their visual content necessarily being similar. In a similar vein, Jörgensen (1995) found that style and type, as art historical attributes, were referenced in names of illustrations. This type of criteria may also be seen in web image queries, as collections (e.g., stock photography) are specified (Jansen, 2008).

Participants explicitly combined several categorization criteria in category names, creating multiclass categories. The share of multiclass categories was 10% for experts with context, 21% for experts without context and 34% for nonexperts. Nonexperts' categories also referenced the most classes overall, on average 1.38 classes per category name. Experts' categories contained 1.23 classes in the no context group and 1.10 classes in the context group. These figures match Jörgensen's (1995), who determined that one third of the image group names were composed of multiple terms (on average, 1.5 terms). Multiclass category names were most frequently constructed when naming categories of photographs of people, further typified with some description of the person(s), photographical attributes, or the story or theme of the photograph. Also Jörgensen found a large share of terms related to people, style and abstract concepts in group names consisting of multiple terms. Half to two thirds of all the terms in the descriptor classes Description, Story, Visual and Photography were found in multiclass category names.

Phase two of the categorization procedure allowed multiple categorizations of the images into the categories constructed in phase one by the participant. In phase two a photograph was assigned to 1.4 categories on average (1.25 for the context group, 1.36 for the no context group and 1.53 for nonexperts). Thus, there was significant overlap in the constructed categories. These overlaps were utilized to create connections in the final model (Section 4.4.5).

4.4.2 Image Similarity Based on Categorization

Nonmetric multidimensional scaling was performed on the similarity data of images without context in order to investigate the dimensions utilized in the similarity evaluations. To characterize the axes Pearson correlation coefficients were calculated between the MDS coordinates of the images on the axes and the frequency of use of the main classes for the images. The results are presented in three dimensions with a stress value of .11 which may be considered good for the number of data points used (Borg & Groenen, 2005). A further increase in dimensionality did not decrease the stress significantly (.08 for 4D, .07 for 5D) while a two-dimensional solution had a higher stress value (.17). This indicates that three dimensions were sufficient to represent the systematic structure in the data. The strongest correlations ($df = 98$) for each axis and their diagonals, all of which are significant at the $p < .001$ level, are visualized in Figures 4 and 5.

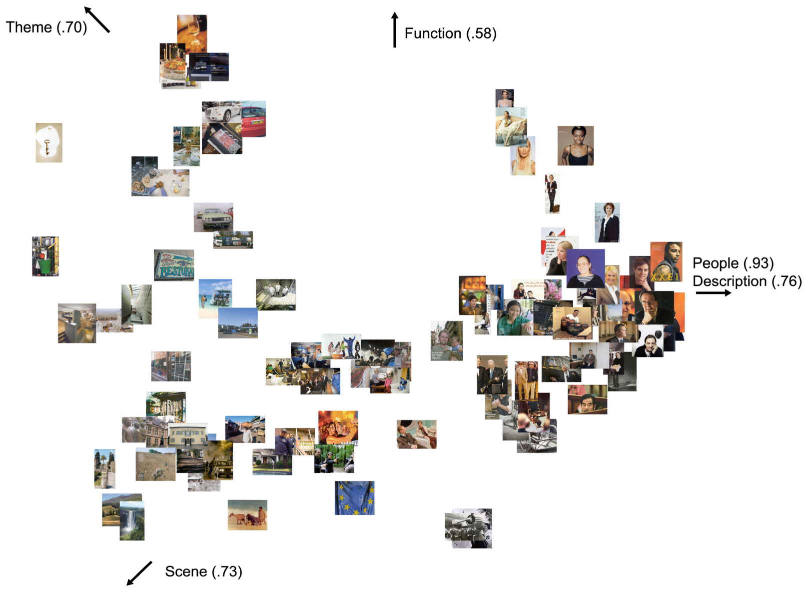

Figure 4. MDS results in dimensions x and y with correlations
Copyright 2010 ASIS&T

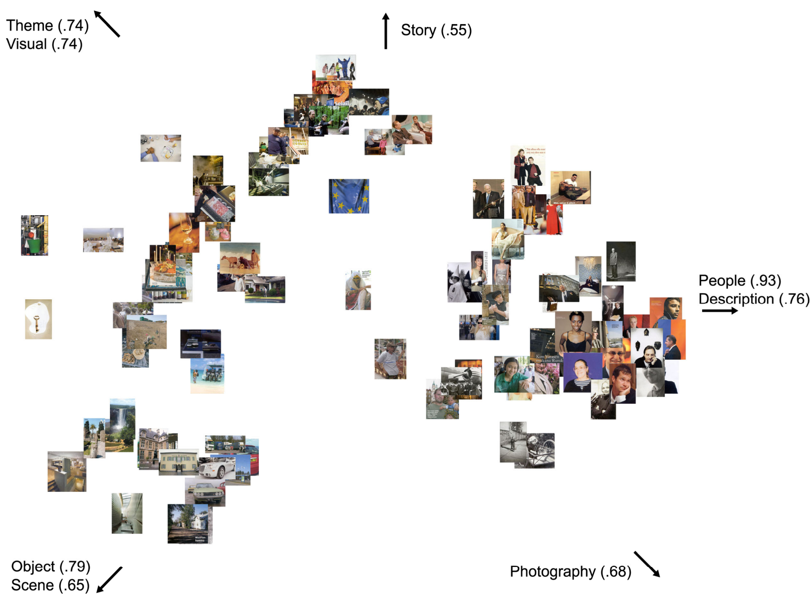

Figure 5. MDS results in dimensions x and z with correlations
Copyright 2010 ASIS&T

The scaling results show emerging axes that divide the image set according to important similarity criteria. The first axis (x axis) divides the set according to the presence of people. The presence of people has been found an important image similarity dimension in past studies (Graham, Friedenberg, Rockmore, & Field, 2010; Rogowitz et al., 1998; Teeselink et al., 2000). Images are also divided into functionally oriented images (e.g., product images or portraits) and shots of more complex scenes (y axis). Photographs of people have previously been found to be sorted according to poses and activities (Rorissa & Hastings, 2004). The third axis further expands on the division between object images and visual storytelling photographs (z axis). Journalists have been shown to consider image background in similarity evaluations (Sormunen et al., 1999). Also related, a division between objects and landscapes has been previously seen in image sorting studies (Teeselink et al., 2000).

Visual features alone did not emerge as axes even though they may be correlated with similarity judgments (Rorissa, Clough & Deselaers, 2008) and be used to evaluate similarity within topical image sets (Sormunen et al., 1999). The results on image similarity assessments support the category name analysis at the main class level as significant correlations for the axes were observed. Also, the similarity dimensions may provide insight into the connections between the visual and thematic categorization of images.

4.4.3 Effect of Context

In categorization each image is assigned to a group of images, creating a link between these images. Thus, image categories may also be seen as relationship attributes (Shatford Layne, 1994) between the images in that category. The relationship attribute also covers associations with textual work. The effect of magazine page context on image categorization was investigated in this dissertation.

Hierarchical cluster analysis was conducted to see whether image similarity was judged similarly with and without page context. Cluster analysis was preformed with the average-linkage method. The cophenetic correlation coefficient value of the clustering solution was .92 for the context group and .81 for the no context group indicating acceptable quality of the solutions. To evaluate whether the clustering of the images was similar in the two groups, Jaccard's similarity coefficient was calculated. The value of Jaccard's coefficient for the categorizations by the two participant groups was .79. It may be concluded that the groups sorted the images in a similar manner. Page context also had no effect on the number of categories constructed, or on task time.

Nevertheless, as previously reported, there was a significant difference in category naming. This result indicates that, while image sorting criteria and behavior were similar, the interpretations of, and the names given to image groups, were affected by the context. The context typically encouraged thematic category naming. Having the page context available also more likely led to categorization on a single facet and to less overlap between categories.

Based on the participants' comments, participants used the text to interpret the content and meaning of the images. The text was utilized to gather more knowledge about the subject of the photograph and the reason that the photograph had been taken and published. Text was seen as something that anchored (Barthes, 1977), explained (Marsh & White, 2003) and elaborated or extended (Martinec & Salway, 2003) the meaning and purpose of the images. The specific type of relationship between the text and image (e.g., their relative status) did not seem to function as a deciding factor in categorization, but rather, the deciding factor was more likely the additional information acquired through the interpretation of the image context at large. One participant called this process "finding keywords", reflecting the terminology of indexing processes. The extreme case of having information used in the categorizations present in the text would be the actual category names appearing in the articles or captions. The question then becomes, for example, could the theme of an image be extracted from the text accompanying it (Barnard & Forsyth, 2001)? The extraction of textual information could be anchored by visual feature analysis, increasing the detection accuracy. This process could also be employed in image retrieval with the article text forming the basis of a query for a matching image.

4.4.4 Categorization Framework and Evaluation

Based on the analysis of the categorization criteria a magazine image categorization framework containing ten main classes (Table 10) was devised. The average class coherence, as evaluated by the Dice coefficient, was .53. Overall, the class coherences were deemed to be acceptable for the inclusion of the classes in the evaluation study (Study two, VI).

Table 10. Main classes of the categorization framework

Main class	Definition
Function	Function served or purpose image was taken for
People	Description of person(s) and their properties
Object	Depicted entities, e.g., inanimate objects or buildings
Scene	Depicted view, e.g., landscape
Theme	Subject matter or field to which content is related
Story	Story conveyed; event or action depicted
Affective	Atmosphere conveyed or emotion invoked in the viewer
Description	Further description of content e.g., adjectives, amounts
Visual	Visual elements, e.g., color, shape, repetition
Photography	Photographic properties, e.g., distance, cropping

In the evaluation study the participants used the framework to categorize another set of images to as many categories as they saw fit. They made an average of 8.2 category selections per image (SD = 3.7). Table 11 displays the share of each main class in the categorizations, the number of times it was used on average per image and participant, and the evaluated usefulness of the class. The main classes Function, Theme and People were the predominant classes in the categorization task: their selections accounted for nearly two thirds of all of the category selections.

Table 11. Use of main classes and average usefulness (scale 1-5)

Main class	Share of categorizations	Average uses per image	Average usefulness
Function	23.8%	2.0±0.4	4.7±0.6
Theme	22.3%	1.8±0.5	4.7±0.6
People	19.0%	1.6±1.5	4.6±0.6
Object	8.7%	0.7±0.4	4.2±0.9
Scene	6.4%	0.5±0.4	4.3±0.9
Photography	6.3%	0.5±0.4	3.6±1.0
Visual	5.7%	0.5±0.2	3.2±1.2
Story	4.5%	0.4±0.3	3.3±1.1
Affective	1.9%	0.2±0.1	2.9±1.1
Description	1.6%	0.1±0.1	2.8±0.9

The average number of categorizations was considerably higher than the average number of categorization criteria in category names in Study one and in the image group names analyzed by Jörgensen (1995). This observation can be attributed to changes in the nature of the task: participants made category selections from a predefined list instead of similarity sorting and category naming. However, despite these changes in the procedure, the main classes had consistent use across the two studies (Figure 6). The use of the main classes in the evaluation study was similar to their use by expert participants categorizing images without context ($\chi^2(8) = 15.35$, $p > .05$).

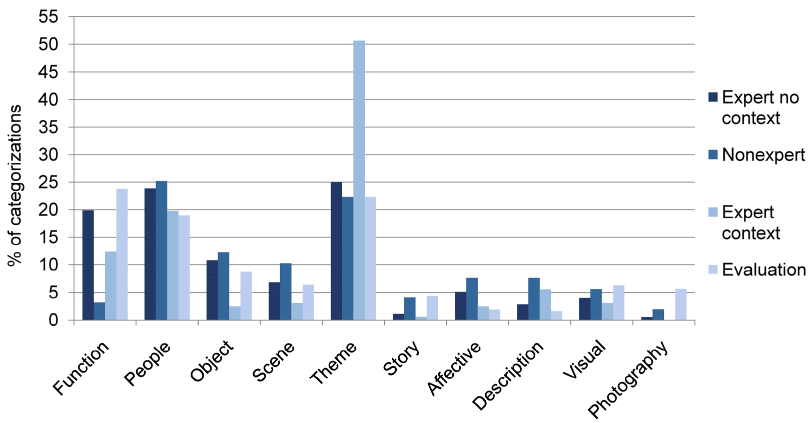

Figure 6. Distributions of classes by participant group

In previous work, the availability of an indexing template slightly changed nonexpert participants' use of image attributes for the same images (Jörgensen, 1996). Jörgensen suggests that such templates may be more useful for expert indexers, and indeed here their use of the categorization criteria stayed consistent.

The average Rolling's measure of classifier agreement across all of the images was .47. Thus, approximately half of any two participants' categories were the same for any of the test images, indicating that categorization was fairly consistent across participants. The participant-evaluated fit of the categorization framework for the images ranged between 2.87 and 4.00 on a scale from 1 to 5. The Rolling's measure and average participant-evaluated fit for the images were correlated ($r(18) = .42$, $p < .05$), thereby suggesting that users could tell when the categorization framework had suitable categories. This result provides a good starting point for further development and modification into specific application contexts.

The framework was evaluated as a useful tool in categorizing images in the given task. Several participants commented that, in an actual work situation, more specific information about the image would have been available (e.g., provenance of the image, information about possible processing, names of persons or specific locations portrayed). This expectation is reasonable, but these types of metadata and indexing terms were outside the scope of this study. It should be noted that the framework may be used on various levels of image semantics depending on the application. For example, a scene may be evaluated as a generic, specific, or abstract setting (Shatford, 1986). The classes used the least and thought to be difficult to use (Story, Affective, Visual, Photography) were mentioned by participants to be “abstract”, “interpretive” or “subjective”. Still, many participants commented that these classes were important and would be employed more for other image sets (e.g., symbolic images) or by other indexers (e.g., photographers who are more knowledgeable of specific terminology).

4.4.5 Final Categorization Model

The final categorization model was created from the classes of categorization criteria and the connections between the classes. The connections were analyzed based on the participants’ strategies for combining them in category names and selections. They further shed light on multifaceted descriptions of images and may be used to develop categorization practices and recommendation tools. For example, if the People class is chosen, a Story class most likely could also be specified.

Figure 7 shows the six strongest symmetrical connections for each dataset, normalized for the use of the classes. Common connections for both studies include Function-Theme, People-Function, People-Theme and Object-Theme. People was also commonly used in conjunction with content descriptor classes in category names, while, in the evaluation study, the class Function was used the most and thus it was often combined with many classes. The results speak for the necessity of specifying the Function (purpose of use) as well as Theme, People and Object (semantic content) in magazine images. While less common in the studies, the content descriptor classes were important for certain image types (e.g., photographs of people) and for certain users. They have also been determined to be important in free description tasks (I; Jörgensen, 1998), connected to image indexing.

This model is of use when evaluating the suitability of image indexing approaches for journalistic images and when developing tools for guiding these processes. These findings may inform the selection of concepts for automated detection, as they provide insight into the types of similarity and categorization criteria that domain experts employ. In VI the model subclasses were mapped to the image metadata schemes from the International Press Telecommunications Council (IPTC), with the conclusion that some classes are currently detected while others pose challenges for automated classification. The categorization model may be taken as a point of comparison and development for concept ontologies (Naphade et al., 2006) in the context of magazine imagery, as these results offer information in terms of the utility and coverage of the classes.

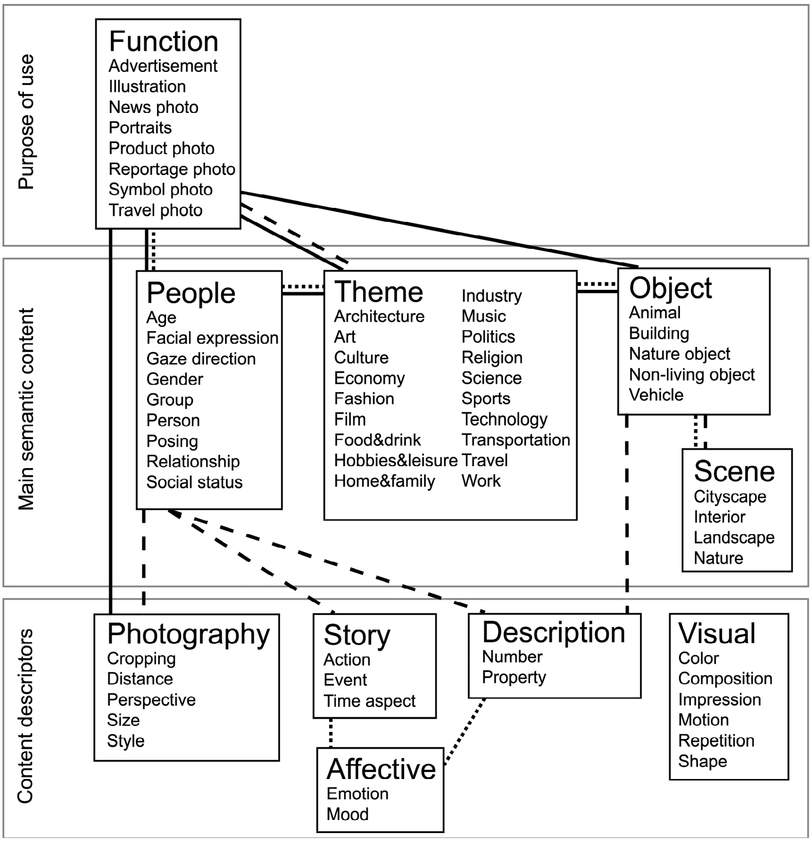

Figure 7. The magazine image categorization model with connections in category names (dashed line) and overlapping categorizations (Study one, dotted line and Study two, solid line) Adapted from VI. Copyright 2010 ASIS&T

4.5 Contributions across Studies

The goal of this dissertation was to understand the processes in journalistic image access and the effects of several types of contextual factors on preferred access points. This work focused on image access in closed collections in a journalistic work context. The results are discussed here across the studies conducted, related to preferred access points for individual images and image categories as well as image searching and categorization behavior. Contextual effects and findings related to domain expertise are also summarized. Finally, an outline for the functionalities of systems that support journalistic image access is presented.

The modeled journalistic image selection process (III) relied on textual queries, image categorization, browsing, image comparisons, and use of textual information to verify relevance. It extends the model of Markkula and Sormunen (2000) by further explicating the browsing and comparison strategies, and the uses of visual and textual information. Text was the main query mode in both III and IV. The textual components of queries were short compared to previous results from other domains. In combined queries the textual component seemed to anchor content-based searches, indicating a specialized multimodal search strategy. The text associated with images was also used in relevance assessments to verify topicality (III) and in image categorization to name thematic categories (III, V).

Certain types of image attributes seem to be preferred in journalistic image access. All image content levels in the image attribute typology (II), i.e., nonvisual, syntactic and semantic, were used in the description of photographs (I). The majority of descriptive terms were semantic but syntactic terms accounted for over 20% of the terms for the free description task, and nearly 10% for the keywording task. Nonvisual attributes were scarce because of the limited context in which the images were presented, making up less than 1% of the terms in both tasks. The developed magazine image categorization model (VI) may also be compared to the image attribute typology (II). The most prevalent categorization criteria belong to classes that correspond to semantic attributes (People, Theme, Object, Scene). The functional aspects important for categorization are not covered in the typology beyond some nonvisual attributes. The content descriptors in the magazine categorization model may be of any type (nonvisual, semantic, syntactic). The criteria used to evaluate magazine image similarity and names assigned to the categories were discovered in V-VI. Image categories have been considered to be dependent on the task, user and environment of the interaction with images (Mojsilovic & Rogowitz, 2001b). The results reported here show however that for experts the distribution of different categorization criteria was unaffected by the participant group, the assigned task or the image set (VI). The categorization model has proved to be a useful tool for analyzing results in other image categorization studies (Laine-Hernandez & Westman, 2010).

Different tasks elicited different types of image descriptions. Descriptions of individual images (I) were mostly generic semantic while names given to image groups (V-VI) had mostly thematic or functional bases. This finding extends the results from Rorissa (2008) who determined that groups of images were labeled using more superordinate level terms while individual image descriptions were mainly at the basic level. Constraining the description of images – either for indexing or searching – was shown to have effects on the types of descriptions gathered. Constraining the description to five keywords resulted in more references to abstract concepts and less visual features than were found in free description (I). Relaying image needs as queries to a search system resulted in more thematic expressions of image needs than requests to intermediaries (III). These observations extend the results of Hollink et al. (2004) who found that free descriptions of images had more perceptual terms than descriptions for query purposes. It appears that constraints (labeling more than one image at a time, using a limited number of terms, matching terms to a retrieval system) make viewers describe their impressions on an abstract, thematic level, while unconstrained description stays more on the level of objects and locations.

Contextual effects on image access were discovered on various levels of context in information retrieval (as defined by Cool & Spink, 2002). There were also differences between experts and nonexperts in the journalistic domain. These two types of effects are summarized in Table 12. Work tasks influenced the interpretation of image needs and were used to create situational relevance criteria (III). The task type affected the queries and search process in multimodal image retrieval (IV). Differences were found in the types of search interactions engaged in by experts and nonexperts (IV). Domain expertise also affected image descriptions as evidenced by differences in category naming (VI). This relationship has implications for the design of image access solutions and user studies.

Table 12. Contextual and expertise effects in the studies

Level of context (Cool & Spink, 2002)	Contextual effect	Expertise effect
Information environment	Work tasks influenced search process and goals (III)	
Information seeking	Image selection based on contextual criteria (III)	
Information retrieval interaction	Task type affected search strategies (IV)	Expertise affected use of query modes and reformulation (IV)
Query or description	Contextual facets in queries (III) Publication context affected image category naming (V)	Expertise affected image category naming (VI)

The studied processes of image description, categorization, and searching need to be supported in journalistic work tasks via image access solutions. Hastings (1999) suggested a framework for the evaluation of image retrieval systems based on the complexity of the retrieval task. In a similar vein, the contributions of this dissertation together with knowledge from reviewed literature may be used to create evaluation criteria for image access systems depending on the access tasks. Journalistic image management and retrieval systems need to account for multifaceted indexing of images based on free descriptions as well as more controlled approaches depending on the type of description required. Multimodal queries by nonvisual, syntactic, and semantic attributes on different levels (generic, specific, abstract) need to be supported. The contextual facets in image requests need to be accounted for and utilized in developing access modes. Browsing functionalities need to be provided for images, image groups and associated textual information together with categorization views. Images should be categorized based both on their content and possible functions. Support for search sessions includes tools for modifying queries and exploring search strategies, collaborative retrieval, and comparing images and making image selections.

5 Conclusions

Journalism deserves consideration as a specific domain for image access in terms of description, categorization and searching. Categorization of journalistic images is based not only on image content but also on functional aspects. Image search tasks and processes in journalistic setting are highly contextual. The contexts of image production, use and retrieval merit more interest in research. The effect of page context on image category naming has implications for the design of post-production indexing solutions. Manipulation of image context and description task could be used to elicit descriptions of images on different levels. The finding that viewers use page context to interpret magazine image content indicates possibilities for image retrieval approaches that employ text mining techniques in conjunction with content-based algorithms to discover the topic of images.

The type of image access required should inform the types of descriptions required and elicited at the image capture, description and organization phases. Knowledge of the categorization criteria can be used to guide research on what types of image information to exploit in retrieval systems as access points for querying and browsing. As a whole, the findings of this dissertation speak of the need for a multifaceted approach to image description, categorization and searching. Multiple access points should be created for journalistic images, corresponding to their content and function. Despite the difficulties in predicting the future utility of images participants grouped images according to how they would function in a use context and later copiously assigned images into functional categories. Typical approaches to describing images might cover the facets of who, what, where and when, but fail to cover the why, which is often a key element in image journalism (Kobré, 2000). Function attributes describe why the image was created or published, an important access point for future uses. A functional approach to image description could aid retrieval through the analysis of image features that correlate with specific image functions. In this sense, image function and other contextual features could serve to span the semantic gap.

For enabling enhanced access to visual content, more effort should be put into combining textual and visual search modes. While content-based retrieval methods may have limited utility on their own, having multiple access strategies to visual content is beneficial. Also, images are published in multimodal contexts and image needs may include multiple facets, combining conceptual and visual criteria. Search interfaces should offer multiple access strategies and support for multimodal query formulation. In the future the description of still images may be cast into the larger issue of retrieval of moving imagery, i.e., video content. In this context the visual nature of the content becomes even more pronounced and visual access modes in both frame-based and temporal manner are needed.

Bibliography

Agosti, M. (2007). *Information Access through Search Engines and Digital Libraries*. New York, NY: Springer-Verlag.

André, P., Cutrell, E., Tan, D. S., & Smith, G. (2009). Designing Novel Image Search Interfaces by Understanding Unique Characteristics and Usage. In T. Gross, J. Gulliksen, P. Kotz, L. Oestreicher, P. Palanque, R. Oliveira Prates & M. Winkcler (Eds.), *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (pp. 340-353). Berlin, Heidelberg: Springer-Verlag,

Armitage, L.H., & Enser, P.G.B. (1997). Analysis of user need in image archives. *Journal of Information Science, 23*(4), 287–289.

Barnard, K., & Forsyth, D. (2001). Learning the Semantics of Words and Pictures. *Proceedings of the IEEE International Conference on Computer Vision, 2,*408-415.

Barthes, R. (1977). Image-Music-Text. London: Fontana.

Bates, M.J. (1979). Information Search Tactics. *Journal of the American Society for Information Science, 30, 205-214.*

Bates, M.J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review 13*(5), 407-424.

Beaudoin, J. (2008). A preliminary investigation of image indexing: The influence of domain knowledge, indexer experience and image characteristics. In J. Lussky, (Ed.), *Proceedings 19th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research.*

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Borg, I., & Groenen, P.J.F. (2005). Modern multidimensional scaling. 2nd edition. New York: Springer.

Borlund, P. (2003). The iIIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research 8*(3).

Burford, B., Briggs, P., & Eakins, J. (2003). A taxonomy of the image: On the classification of content for image retrieval. *Visual Communications, 2*(2), 123-161.

Chalmers, M. (1999). Comparing Information Access Approaches. *Journal of the American Society for Information Science, 50*(12), 1108–1118.

Chen, H. (2001). An Analysis of Image Queries in the Field of Art History. *Journal of the American Society for Information Science and Technology, 52*(3), 260-273.

Chen, H., & Rasmussen, E.M. (1999). Intellectual Access to Images. *Library Trends, 48*(2), 291-302

Choi, Y. (2009). Effects of contextual factors on image searching on the Web. *Journal of the American Society for Information Science and Technology 61*(10):2011-2028.

Choi, Y., & Rasmussen, E. (2002). Users' relevance criteria in image retrieval in American history. *Information Processing and Management, 38*(5), 695-726.

Choi, Y., & Rasmussen, E. M. (2003). Searching for images: The analysis of users' queries for image retrieval in American history. *Journal of the American Society for Information Science and Technology, 54*(6), 498–511.

Chung, E., & Yoon, J. (2011). Image needs in the context of image use: An exploratory study. *Journal of Information Science,* 37(2):163-177.

Conniss, L. R., Ashford, A. J., & Graham, M. E. (2000). *Information seeking behavior in image retrieval: VISOR I Final report*. Newcastle: Institute for Image Data Research.

Cool, C., & Spink, A. (2002). Issues of context in information retrieval (IR): an introduction to the special issue. *Information Processing & Management, 38*(5), 605-611.

Cunningham, S. J., & Masoodian, M. (2006). Looking for a Picture: An Analysis of Everyday Image Information Searching. *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries (JCDL '06)*, 198-199.

Datta, R., Joshi, D., Li, J., & Wang, J.Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2), 1-60.

Eakins, J.P., & Graham, M. E. (1999). *Content-based Image Retrieval: A report to the JISC Technology Applications Programme*. Retrieved from http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc

Eakins, J.P., Briggs, P,. & Burford, B. (2004). Image Retrieval Interfaces: A User Perspective. *Proceedings of CIVR 2004, Lecture Notes in Computer Science, 3115,* 628-637.

Enser, P.G.B (2000). Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science, 26*(4), 199-210.

Enser, P.G.B. & Sandom, C.J. (2002). Retrieval of archival moving imagery - CBIR outside the frame*? Proceedings of CIVR 2002, Lecture Notes in Computer Science, 2383*, 206-214.

Enser, P.G.B. (1993). Query analysis in a visual information retrieval context. *Journal of Document and Text Management, 1*(1), 25-52.

Enser, P.G.B., & Sandom, C.J. (2001). Retrieval of archival moving imagery: a step too far for CBIR? In: N. Boujemaa, C.-S. Li, & D. Forsyth (Eds.), *Proceedings of MMCBIR' 2001: Multimedia Content-Based Indexing and Retrieval Workshop*, 7-10.

Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1), 61-74.

Fidel, R. (1997). The image retrieval task: implications for the design and evaluation of image databases. *The New Review of Hypermedia and Multimedia, 3*, 181-199.

Frost, C. (2001). The role of mental models in a multi-modal image search. *Proceedings of the American Society for Information Science and Technology, 38,* 52-57.

Frost, C.O., Taylor, B., Noakes, A., Markel, S., Torres, D., & Drabenstott, K.M. (2000). Browse and Search Patterns in a Digital Image Database. *Information Retrieval, 1*(4), 287-313.

Garber, S.R., & Grunes, M.B. (1992). The art of search: A study of art directors. *Proceedings of the Conference on Human Factors in Computing Systems (CHI '92)*, 157–173.

Goodrum, A. A., Bejune, M. M., & Siochi, A. C. (2003). A State Transition Analysis of Image Search Patterns on the Web. *Proceedings of CIVR 2003, Lecture Notes in Computer Science, 2728,* 281-290.

Goodrum, A., & Spink, A. (2001). Image searching on the Excite Web search engine. *Information Processing and Management, 37*(2), 295-311.

Graham, D.J., Friedenberg, J.D., Rockmore, D.N., & Field, D. J. (2010). Mapping the similarity space of paintings: Image statistics and visual perception. *Visual Cognition 18*(4), 559-573.

Greisdorf, H. & O'Connor, B. (2002b). What do users see? Exploring the cognitive nature of functional image retrieval. *Proceedings of the American Society for Information Science and Technology, 39*, 383–390.

Greisdorf, H., & O'Connor, B. (2002a). Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation, 58*(1), 6-29.

Hall, S. (1981). The determinations of news photographs. In S. Cohen & J. Young (Eds.) *The Manufacture of News: Social Problems, Deviance and the Mass Media* (pp.226-243). London, UK: Constable.

Hasler, D., & Süsstrunk, S.E. (2003). Measuring colourfulness in natural images. *Proceedings of IS&T/SPIE Human Vision and Electronic Imaging VIII, 5007,* 87-95.

Hastings, S.K. (1995). Query categories in a study of intellectual access to digitized art images. In T. Kinney (Ed.), *Proceedings of the 58th Annual Meeting of the American Society for Information Science* (pp. 3–8). Medford, NJ: Information Today.

Hastings, S.K. (1999). Evaluation of image retrieval systems: Role of user feedback. *Library Trends, 48*(2), 438–452.

Heidorn, P.B. (1999). Image Retrieval as Linguistic and Nonlinguistic Visual Model Matching. *Library Trends, 48*(2), 303-325.

Hollink, L., Schreiber, G., Wielinga, B., & Worring, M. (2004). Classification of User Image Descriptions. *International Journal of Human-Computer Studies, 61*(5), 601-626.

Hung, T-Y. (2005). Search Moves and Tactics for Image Retrieval in the Field of Journalism: A Pilot Study. *Journal of Educational Media and Library Sciences, 42*(3), 329-346.

Ingwersen, P. (1992). *Information Retrieval Interaction*. London: Taylor Graham.

Ingwersen, P., & Järvelin, K. (2005). *The Turn. Integration of Information Seeking and Retrieval in Context*. Berlin: Springer.

International Press Telecommunications Council IPTC (2008a). *Photo metadata specification*. Retrieved from http://www.iptc.org/std/photometadata/2008/specification/IPTC-PhotoMetadata-2008.pdf

International Press Telecommunications Council IPTC (2008b). *Descriptive NewsCodes*. Retrieved from http://www.iptc.org/cms/site/index.html?channel=CH0103#descrncd

Jacob, E.K. (2004). Classification and categorization: a difference that makes a difference. *Library Trends, 52*(3), 522.

Jaimes, A. & Chang, S-F. (2000). A Conceptual Framework for Indexing Visual Information at Multiple Levels. *Proceedings of IS&T/SPIE Internet Imaging I, 3964,* 2-15.

Jaimes, A. (2006). Human factors in automatic image retrieval system design and evaluation. *Proceedings of IS&T/SPIE Internet Imaging VII, 6061,* 101-109.

Jansen, B.J. (2008). Searching for digital images on the web. *Journal of Documentation, 64*(1), 81-101.

Jespersen, H.P., & Jespersen, J.K. (2004). The problem of subject access to visual materials. *Journal of Educational Media and Library Science, 42*(1), 37-48.

Jörgensen, C. (1995). Classifying images: criteria for grouping as revealed in a sorting task. In R. Schwartz (Ed.), *Proceedings of the 6th ASIS SIG/CR Classification Research Workshop 32* (pp. 45-63).

Jörgensen, C. (1996). Testing an image description template. *Proceedings of the American Society for Information Science, 43,* 209-213.

Jörgensen, C. (1998). Attributes of images in describing tasks. *Information Processing & Management, 34*(2), 161-174.

Jörgensen, C. (1999). Access to Pictorial Material: A Review of Current Research and Future Prospects. *Computers and the Humanities, 33*(4), 293-318.

Jörgensen, C. (2001). Introduction and overview. *Journal of the American Society for Information Science and Technology, 52*(11), 906-910.

Jörgensen, C., & Jörgensen, P. (2005). Image querying by image professionals. *Journal of the American Society of Information Science and Technology, 56*(12), 1346-1359.

Keister, L.H. (1994). User Types and Queries: Impact on Image Access Systems. In R. Fidel, T. B. Hahn, E. M. Rasmussen & P. J. Smith (Eds.), *Challenges in Indexing Electronic Text and Images* (pp. 7-22). Medford, NJ: Learned Information, Inc.

Kobré, K. (2000). *Photojournalism: The Professionals' Approach*. Boston, MA: Focal Press.

Laine-Hernandez, M. & Westman, S. (2008). Multifaceted Image similarity criteria as revealed by sorting tasks. *Proceedings of the American Society for Information Science and Technology, 45*, 1-14.

Laine-Hernandez, M., & Westman, S. (2010). Comparison of categorization criteria across image genres. *Proceedings of the American Society for Information Science and Technology, 47,* 1–12.

Lancaster, F.W. (1998). *Indexing and abstracting in theory and practice*. Champaign-Urbana: GSLIS.

Lee, H.-J., & Neal, D. (2010). A new model for semantic photograph description combining basic levels and user-assigned descriptors. *Journal of Information Science, 36*(5), 547-565.

Lohse, G.L., Biolsi, K., Walker, N., & Rueter, H.H. (1994). A classification of visual representations. *Communications of the ACM, 37*(12), 36-49.

Lohse, J., Rueter, H., Biolsi, K., & Walker, N. (1990). Classifying visual knowledge representations: a foundation for visualization research. *Proceedings of IEEE Visualization*, 131-138.

Markey, K. (1984). Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library & Information Science Research, 6*(2), 155-177.

Markkula, M., & Sormunen, E. (1998). Searching for Photos - Journalists' Practices in Pictorial IR. In J. Eakins, D. Harper & J. Jose (Eds.), *The Challenge of Image Retrieval.* Electronic Workshops in Computing (eWIC) Newcastle upon Tyne, UK.

Markkula, M., & Sormunen, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo archive. *Information Retrieval, 1*(4), 259-285.

Marsh, E.E., & White, M.D. (2003). A taxonomy of relationships between images and text. *Journal of Documentation 59*(6), 647-672.

Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication 4*(3), 337-371

McCay-Peet, L., & Toms, E. (2009). Image use within the work task model: Images as information and illustration. J*ournal of the American Society for Information Science and Technology, 60*(12), 2416–2429.

McDonald, S., & Tait, J. (2003). Search strategies in content-based image retrieval. *Proceedings of SIGIR,* 80-87.

Mojsilovic, A., & Rogowitz, B. (2001a). A Psychophysical approach to modeling image semantics. *Proceedings of SPIE Human Vision and Electronic Imaging VI, 4299,* 470-477.

Mojsilovic, A., & Rogowitz, B. (2001b). Capturing Image Semantics with Low-level Descriptors. *Proceedings of IEEE International Conference on Image Processing, ICIP 2001, 1*, 18-21.

Naphade, M., Curtis, J., Hauptmann, A., Kennedy, L., Hsu, W., Chang, S-F., & Smith, J.R. (2006). Large-scale concept ontology for multimedia. *IEEE Multimedia, 13*(3), 86-91.

Neal, D. (2008). News Photographers, Librarians, Tags, and Controlled Vocabularies: Balancing the Forces. *Journal of Library Metadata, 8*(3), 199-219.

Ørnager, S (1995). The newspaper image database: empirical supported analysis of users' typology and word association clusters. In E.A. Fox, P. Ingwersen & R. Fidel (Eds.), Proceedings of the 18th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval (pp. 212–218).

Ørnager, S. (1997). Image retrieval: Theoretical analysis and empirical user studies on accessing information in images. *Proceedings of the American Society for Information Science and Technology, 34*. 202-211.

Pettersson, R. (1998). Image functions in information design. In R. E. Griffin, W. J. Gibbs, & B. Wiegman (Eds.*), Visual Literacy In An Information Age* (pp. 259-268). International Visual Literacy Association.

Rasmussen, E. M. (1997). Indexing images. *Annual Review of Information Science and Technology, 32*, 169-196.

Rogowitz, B.E., Frese, T., Smith, J.R., Bouman, C.A. & Kalin, E. (1998). Perceptual image similarity experiments. *Proceedings of SPIE Human Vision and Electronic Imaging III, 3299,* 576-590.

Rolling, L. (1981). Indexing consistency, quality and efficiency. *Information Processing and Management, 17*(2), 69-76.

Rorissa, A. (2008). User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory. *Information Processing and Management, 44*(5), 1741-1753.

Rorissa, A. (2010). A comparative study of flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology, 61*(11), 2230 - 2242.

Rorissa, A., & Hastings, S.K. (2004). Free sorting of images: Attributes used for categorization *Proceedings of the American Society for Information Science and Technology, 41*, 360-366.

Rorissa, A., & Iyer, H. (2008). Theories of cognition and image categorization: What category labels reveal about basic level theory. *Journal of the American Society for Information Science and Technology, 59*(9), 1383–1392.

Rorissa, A., Clough, P., & Deselaers, T. (2008). Exploring the relationship between feature and perceptual visual spaces. *Journal of the American Society for Information Science and Technology, 59*(5), 770–784.

Rowley, J., & Hartley, R. (2008). *Organizing knowledge. An introduction to Managing Access to Information*. Aldershot: Ashgate Publishing.

Santini, S., Gupta, A., & Jain, R. (2001). Emergent Semantics through Interaction in Image Databases. *IEEE Transactions on Knowledge and Data Engineering, 13*(3), 337-51.

Saracevic, T. (1997). The stratified model of information retrieval interaction: extension and applications. *Proceedings of the American Society for Information Science and Technology, 34*, 313-327.

Saunders, S. (2010). *IPTC Controlled Vocabulary*. Presentation at the Fourth international Photo Metadata Conference, IPTC. Retrieved from http://www.phmdc.org/phmdc2010/PhMdC2010--SSaunders-ImageCV.pdf

Shatford Layne, S. (1994). Some issues in the indexing of images. *Journal of the American Society for Information Science, 45*(8), 583-588.

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly, 6*(3), 39-62.

Siochi, A. C., & Ehrich, R. W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. *ACM Transactions on Information Systems, 9*(4), 309-335.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(*12), 1349–1380.

Sonnenwald, D.H. (1999). Evolving perspectives of human behavior: contexts, situation, social networks and information horizons. In: T. Wilson & D.K. Allen (Eds.), *Exploring the contexts of information behaviour: Proceedings of the 2nd international conference on research in information needs, seeking and use in different contexts* (pp. 176–190). London, UK: Taylor Graham.

Sormunen, E., Markkula, M., & Järvelin, K. (1999). The Perceived Similarity of Photos - A Test-Collection Based Evaluation Framework for the Content-Based Photo Retrieval Algorithms. In: S.W. Draper, M.D. Dunlop, I. Ruthven & C.J. van Rijsbergen (Eds.), *Mira 99: Evaluating interactive information retrieval.*

Spink, A., Goodrum, A., & Robins, D. (1998). Elicitation Behaviour During Mediated Information Retrieval. *Information Processing and Management, 34*(2-3), 257-273.

Stvilia, B., & Jörgensen, C. (2009). User-generated collection level metadata in an online photo-sharing system. *Library & Information Science Research. 31*(1), 54-65.

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology, 61*(12), 2477-2489.

Taylor, A.G., & Joudrey, D. N. (2009). *The organization of information.* Westport:Libraries Unlimited.

Teeselink, I.K., Blommaert, F., & de Ridder, H. (2000). Image Categorization. *Journal of Imaging Science and Technology, 44*(6), 552-559.

Tjondronegoro, D., Spink, A., & Jansen, B.J. (2009). A study and comparison of multimedia Web searching: 1997–2006. *Journal of the American Society for Information Science and Technology, 60*(9), 1756-1768.

Urban, J., & Jose, J. M. (2006). Evaluating a workspace's usefulness for image retrieval. *Multimedia Systems, 12*(4-5), 355-373.

Vailaya, A., Jain, A., & Zhang, H.J. (1998). On Image Classification: City Images vs. Landscapes. *Pattern Recognition, 31*(12), 1921-1935.

van Rijsbergen, C.J. (1979). *Information Retrieval.* London: Butterworths.

Wilson, T.D. (1999). Exploring models of information behavior: the uncertainty project. *Information Processing & Management*, 35(6), 839-849.

Errata

1. In II, a cell Table 4, in the Attribute type column was empty. The missing text (Specific) may be seen in Table 4 in this dissertation.

2. In III, the percentages of three generic facets in Table 2 were incorrectly reported. An updated version of the publication with corrected values is available as a supplemental file at http://doi.acm.org/10.1145/1164820.1164843#supp

3. In V, the image similarity measure P should be P = p(i, j)/p(i).

4. In VI, there was a connection missing from Figure 8. The figure may be found in its correct form as Figure 7 in this dissertation.

Aalto-DD 87/2011

ISBN 978-952-60-4293-0 (pdf)
ISBN 978-952-60-4292-3
ISSN-L 1799-4934
ISSN 1799-4942 (pdf)
ISSN 1799-4934

Aalto University
School of Science
Department of Media Technology
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS