Smashing Reality (in)to Bits?

Views on representation and interpretation in Image Retrieval

1. Introduction

Pictures have the capability of conveying much information in a limited amount of space. In this sense, the marketing advertisement from 1927 by Fred R. Barnard stating: "One Picture is Worth Ten Thousand Words", makes sense even though it has turned into a cliché.

The inherent ability pictures have to get across information makes them very useful for documentation purposes in many areas. Domains working with satellite surveillance, medicine, history, media, entertainment, geology, biometrics and astronomy, are some of the professional fields actively using representations of pictures in their work.

One aspect that makes pictures so powerful, important and "worth" so much as providers of information is that virtually no form of textual description of an object or situation will ever add up to a depiction of it. According to Mitchell (1994), this is because a text-based representation cannot represent an object in the same way a visual representation can:

> *..It may refer to an object, describe it, invoke it, but it can never bring its visual presence before us in the way pictures do* (p 152).

Thus, much of the value in using images lies in the fact that many aspects of particular situations/events are quite challenging to put into writing. One example may be how to describe a cover of clouds if being unfamiliar with the terminology[1]. A different example is how to describe images of a hurricane as it moves.

However, an implicit assumption underlying the points of view presented above seems to be that a human being is always involved in the interpretation/understanding/description of the depicted content.

In this sense, a picture may certainly be worth ten thousand words, but at the same time the worth of a picture in turn relies on human intelligence, knowledge, experience and intuition. This may pose quite a challenge in situations where people are not directly involved in the process of perceiving and describing pictures.

In order to be able to store, manage, and retrieve images to be used in a computer environment, creating representations of real-world objects[2] is a crucial first step. The most direct way of initiating this process is to use a modern digital camera where a representation of the object viewed through the lens (or on a screen) is captured immediately by pressing the shutter button. If the object is already recorded on a physical medium, e.g. a paining or photography, the digital representation may be created using some form of scanning technology.

[1] For instance: cumulus, stratus, cirrus or nimbus clouds are separate cloud types appearing in high, mid, or low-level positions.

[2] Real-world object refers here to all entities (both man-made and natural) possible to record and transform into digital representations.

After capturing a real-world object or picture in a digital format using a camera or a scanner, the digital representation has to be further processed in order to be used for retrieval purposes. This of course is because a computer initially only "knows" the picture to be a blob[3].

Thus, in order to be able to use the contents of a blob for retrieval purposes it has to be further transformed into a specialised digital image representation. These image representations have primarily been created using either some form of explanatory text in the form of keywords or descriptions, or by using the syntactical image content itself.

Extensive surveys done by Rui et al. (1999) and Lev et al. (2006) suggest that, besides the use of text, the most common approach to image retrieval when utilizing the image content has been extraction of the colours, textures, and shapes present in the image.

The surveys by Rui et al. (1999) and Lev et al. (2006) portray previous research efforts, particularly in research on image retrieval relying on the image content, as being focused primarily on assessing which of the features, or combination of features, have been most effective to use for image retrieval purposes. However, neither of the two surveys, nor other available literature[4], raise the questions if, how and to what extent the current approaches to the process of creating image representations actually provide accurate means of interpreting and describing the real-world content depicted in the original picture.

From this, a discussion focusing on if and to what extent the line of thought underlying each of the current techniques for creating image representations actually is suitable may be regarded as paramount. However, relatively little attention seems to have been put on this discussion in the field of image retrieval.

1.1. Essay Purpose and Goal

The main purpose of this essay is to first present briefly some notions of the concepts representation and interpretation, which here are regarded as two central concepts in the creation of digital image representations, and then discuss how various lines of thought addressing these concepts may be related to the two main approaches to the process of creating digital image representations to be used for image retrieval purposes.

Creating image representations that fully describe the depicted content on a level usable for a general purpose retrieval system remains a serious challenge, and in this essay an approach aimed at hopefully alleviating this problem is outlined.

2. Theoretical Foundation

2.1. Representation

According to Mitchell (1995), the term *representation* has since antiquity been a fundamental concept in both aesthetics and semiotics, and the term may in this sense be seen as a triangular relationship between the object represented, how it is represented, and an observer[5]. In

[3] Binary Large OBject

[4] This is not to say that no one has raised the issue previously (as can be seen in Dervin (1977), Neill (1987) and Eliade (1991) where somewhat related issues are discussed), but from the literature available to me at this point, it does not seem to have been a highly prioritised research area in later years.

[5] In which a representation is always being *of* something or someone, *by* something or someone, *to* someone (Mitchell 1995:12).

addition, it may in many cases also be fruitful to include a fourth dimension represented by the creator of the representation.

Central to the view where all four dimensions are present is the humans in the opposite ends of the communication channel facilitated by the representation (pp 11,12). Here, the underlying assumption of having humans present "in the loop" is made explicit.

To Aristotle, representations differed from one another in *object*, *manner*, and *means*[6]. The object is that which is represented, the manner is the way in which the object is represented, and the means is the material used to represent the object. In addition, there is a one-to-many relationship between the means and the manner of a representation where means may be employed in different ways (ibid, 13).

Different Kinds of Visual Representations

Reality may be represented visually in various ways. Here, three different kinds of visual representations are of special interest.

On the most basic level, representations may appear in people's brains in the form of an *Imagination*[7] occurring as a figment of our brain processes, thus only available as visual representations to the person having the imagination. On the other hand, representations may also be completely tangible and available for all to perceive. These representations, for instance in the form of *Pictures*[8], can be converted/transformed into digital representations in the form of *Images*[9]. Images may of course also be created directly or instantaneously, for instance by capturing an object using a digital camera.

Distinguishing between pictures and images in this way resembles the approach taken in Mitchell (1994:4), and helps provide a clear view of the difference between various levels of visual "manifestation" of objects.

2.2. Interpretation

When reading both written material and images there is often a gap between the author and the reader. The gap may of course often refer to distance between author and reader in time and space, but more importantly it may refer to the potential discrepancy between what actually is communicated from the author versus how the communicated message is understood by a reader. The process of interpretation is thus very important.

2.2.1. Interpreting Pictures and Images

Taking in the contents of a picture using our visual senses by looking only at the different patterns of light, gives us a neutral perception of the different elements in the picture. In this case the *percept* refers to what our senses perceive without assistance from any higher mental processes. A *concept* on the other hand, refers to a representation, an abstract or generic idea, generalized from particular instances. In the human mind percept and concept are very tightly coupled. This implies that human interpretation of a picture most often will automatically be

[6] "Means" is referred to as "codes" in Mitchell (1995), representing for instance language, musical forms, paint etc.

[7] The act or power of forming an iconic mental representation of something not present to the senses or never before wholly perceived in reality Merriam Webster.

[8] A design or representation made by various means (as painting, drawing, or photography) Merriam Webster.

[9] Images are a visual representation of a picture (or real-world object) produced on an electronic display (as a television or computer screen) Merriam Webster.

supported by background knowledge in addition to the perception of the content (Jaimes and Chang 2002).

One early approach to the process of creating verbal representations of visual representations is known as *Ekphrasis* (Mitchell 1994:152). This approach to the process of (interpreting and) describing visual imagery has been around since antiquity, and one view is to see it as a literary description of or commentary on a visual work of art, things, a person, or an experience.

A different approach to the interpretation of images arose with the *text analysis* introduced by the Suisse linguist Ferdinand de Saussure, where the term "text" was broadened to also cover pictures (Østbye, Helland et al. 2002). The semiotics of Saussure had one expression-side and one content-side where the expression-side relates to the sign while the content-side relates to the actual meaning of it. The relationship between the two is built on (socially agreed upon) conventions. In semiotics, *denotation* and *connotation* are also central terms. Denotation refers to the literal meaning of an expression, while connotation refers to an additional "deeper" meaning springing into action when the literal meaning is insufficient to fully explain something. Connotations are cultural dependent and thus different between different people (Østbye, Helland et al. 2002).

A concrete, general, and more recent guide to the process of interpreting pictures and images is found in Hilligoss and Howard (2002, pp 34-38). The authors present some general points or categories that may be used when interpreting the contents of pictures, listed here:

- General purpose and appearance
- Overall design
- People in the image
- Image Setting
- Symbols and signs
- Colour features
- Text present
- Image Story

Under each point there is further specified what kind of information that could be relevant to record. From this list we can see that there is a focus on both the syntactical and semantic aspects of the image, underlining the importance of both levels.

2.2.2. A Hermeneutic Approach to Interpretation of Images

According to Hans-Georg Gadamer, hermeneutics should be considered as an approach to the process of interpretation.

In Gadamer's view, the particular meaning a picture has to someone comes from the reflection on, or interpretation of, the depicted content. Thus, to Gadamer, understanding is to always interpret, and it is only through this productive process of interpretation the meaning of the picture can be understood (Gadamer 1975:358).

As can be gathered from the brief presentation above, *understanding* is a central concept with regards to both interpretation and representation. Following from this is the fundamental part of both hermeneutics and Gadamers theory, namely that parts are to be interpreted/understood from the whole, and the whole is to be interpreted/understood from the parts (Gadamer 2003:33). Another central point here is that the "whole" (and hence the understanding of it) consists of both a subjective and objective character (ibid, 34).

When a text/image is analysed according to the hermeneutic line of thought, the analyst is guided by cultural and historical factors. These factors are in the hermeneutic point of view regarded as a form of prejudice, and in this view they are the very reason why humans are being able to interpret the text/image (Østbye, Helland et al. 2002).

3. Representations in Digital Image Collections

Much of the current work in the field of image retrieval is focused around two different approaches with regards to how images are represented.

One approach is to use various kinds of metadata to identify and represent images. This approach may be used to tag or annotate images with keywords based on extracting some of the metadata associated with the image. Using this kind of metadata, like date, time, or GPS location to describe images in this manner is commonly used in approaches focusing on image *context*.

A different approach is to use the depicted image content in order to identify and represent images. This latter approach may be convenient in order to create a representation reflecting the depicted content. The representation may be based on a human interpretation and description of the depicted content, but the description may also be the result of extracting a collection of the visual features to be used to support users in retrieving images. In this essay it is the approaches focusing on image content that are of interest.

The main purpose of *content-focused* [10] image representations, ordinarily consisting either of text provided by a human, or as a collection of image features automatically extracted from the image, is to identify key elements in an image. In this essay, these image representations are thus regarded as the *signatures* of images stored in the collection.

Most of the previous approaches to image retrieval have taken one of these two different paths where one focuses more on the value of semantics while the other focuses more on technical aspects. Some main characters of these approaches are briefly presented in the following.

3.1. The Text Based Approach to Image Retrieval

When creating image representations using text, a common approach in image retrieval is to have people manually describe image contents using text in a way that they feel describe important aspects associated with the depicted content and/or the surroundings of an image. These descriptions are thought to reflect the semantic[11] contents of an image. Thus, human analysis and interpretation of pictures plays a crucial role in fully understanding the semantics of the image contents.

[10] Content-focused is used to separate these approaches form approaches focusing strictly on the use of automatically generated metadata, e.g. like those found in the EXIF-file of an image, but the term is also used to being able to include both the text-based approach focusing on image content and the low-level syntactical approach, which is most often referred to as an "content-based" approach.

[11] The meaning or relationship of meanings of a sign or set of signs; especially: connotative meaning Merriam Webster.

The first image retrieval system combining this kind of text-based image representations with database technology to support image retrieval appeared in the mid 1980s (Prasad, Gupta et al. 1987). This approach is known as *Text Based Image Retrieval* (TBIR) (Chang and Hsu 1992; Baeza-Yates and Ribeiro-Neto 1999; Lu 1999) and is still widely used.

The traditional way of processing images to be stored in digital collections may to some degree be compared to the act of ekphrasis in that the keywords/annotations[12]/descriptions used to represent (and index and retrieve) images often are based on a literal "description" of the image content.

However, image representations created in this manner are not directly connected with the image contents as such as they are the result of an interpretation of the content, and may very well be a figment of the annotator's imagination. A potentially problematic aspect in this sense is thus related to human perception subjectivity (Rui, Huang et al. 1998), and this subjectivity may, according to the critics of the TBIR approach, cause some difficulties when it comes to image retrieval.

Creating image representations using text may be seen as a form of what Fjelland (1999:185) refers to as *synthetic reductionism* in that the keywords/annotations/descriptions resulting from a human interpretation of the image content are creating a "whole" by putting pieces together into some form of text-based description.

However, considering the large degree of human involvement when creating image representations using text in the TBIR approach, and the value put on subjective interpretations, this clearly parts way with for instance a positivistic view of an objective truth.

3.2. The Content Based Approach to Image Retrieval

During the 1990s, *Content Based Image Retrieval* (CBIR) emerged from Computer Vision and Pattern Recognition research as an alternative to text-based image retrieval. Here, image representations are created using the global and/or local low-level syntactical image features (such as colours, textures and shapes) of an image (Rui, Huang et al. 1999). The features are used to index and retrieve images based on a comparison of their visual similarity to a given set of image features compiled in a vector, or collection of vectors.

Compared to the act of ekphrasis, the CBIR approach to creating image representations has circumvented the need for involving a human at run-time in the process in favour of a standard "objective interpretation" of each image based on the steps of a specially developed algorithm of some sort. In a general purpose retrieval system[13], the same algorithm is used on all images fed to the system regardless of the depicted content. The way in which image content is to be interpreted is thus determined by the developer of the algorithm.

The creation of image representations based on the depicted content may be seen as a form of what Fjelland (1999:185) refers to as *analytical reductionism* in that the algorithms are splitting the image content up into collections of features used for indexing and retrieval purposes, hence moving from the whole to the parts.

[12] Compared to keywords, annotation is here seen as a more stringent way of describing the content of image/multimedia contents, e.g. based on some sort of reference model like Dublin Core, parts of MPEG7, or CIDOC CRM etc.

[13] Examples are systems using Oracle (up to v. 10g), IBM (up to v. 8.2), or the LIRE approach.

As CBIR systems are designed to retrieve images similar to an image operating as the search criterion, these systems may for instance aid users in finding both different types of clouds, or being able to track the movement of a hurricane. Another possibility in this kind of systems may be to locate "beautiful" images. This is of course based on the notion that the image provided by the user is depicting clouds, hurricanes, or is an example of what constitutes a beautiful image to the user.

3.2.1. Challenges Associated With Image Representations in CBIR

Known problems in the field of content-based image retrieval, whose occurrence may be attributed to some these problematic assumptions, are often referred to as different types of "gaps".

The limitation associated with the creation of image representations lies in that one cannot determine how an image looks by using the image signature alone as this may be a vector (or set of vectors) presenting only an arithmetic mean of the values to the syntactical features extracted from an image. This deficiency when referring to image features is also called a *numerical Gap*, which refers to weaknesses associated with image signatures consisting of image features in that they have a lack of fidelity to the visual content they represent[14].

The numerical gap reveals itself in the retrieval phase in what is called the semantic gap. The semantic gap refers to the discrepancy that exists between the information currently possible to extract from visual data and the interpretation the same data has for a user in a given situation (Smeulders et. al in Dorai and Venkatesh 2003).

A central point from Dreyfus et al. (1986) applicable here, is to articulate the distinction between "know-how" and "know-that" (Dreyfus, Dreyfus et al. 1986), also often referred to as procedural knowledge and tacit knowledge. Humans may acquire both, but computers are not able to develop and make use of tacit knowledge. This dichotomy results in quite a challenge with regards to computer generated representations to be used for image retrieval as it is the meaning of an image that may be of central value to a user looking for it.

3.3. The Science Underlying TBIR and CBIR

Both the TBIR approach and the CBIR approach are to various degrees rooted in the field of Computer Science, defined as:

> *[…] the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application. The fundamental question underlying all of computing is, "what can be (effectively) automated?"* (Denning, Comer et al. 1989).

By the definition stated above, research in the field of Computer Science may be seen as being driven, or at least heavily influenced, by a form of technological determinism reflecting the notion that society has to follow the technical development because the technical development is important to it. The definition also seems to suggest that research and development should be guided by what Fjelland (1999) refers to as a *technological imperative* stating that whatever may be realised using technology should be realised (p 225).

[14] http://ralyx.inria.fr/2006/Raweb/imedia/uid18.html

According to Timothy Colburn (2004), important aspects in computer science are focused at:

> *[..] theories for understanding computer systems and methods; design methodologies, algorithms and tools; methods for the testing of computer related concepts; methods of analysis and verification of such concepts, as well as tools and methods for knowledge representation and implementation.*

However, this second aspect seems to have been somewhat downplayed both in the fields of Computer Science and image retrieval. In many cases scientists in these fields have in stead focused on design and evaluation, construction of algorithms, and the testing of prototypes to be used in clearly defined settings. Hence, in much of the work taking this approach to image retrieval, the research aspect is overshadowed by the development process. Hence, the end result is often not a general or specific theory, but rather algorithms to be put to use in a computer program or information system. There is thus a quite blurry distinction between theory and practise in these fields. Following from this is that the focus is on the technical aspect, not so much on the human aspect.

Another interesting point in the statement from Colburn on what is important in Computer Science is that the human dimension seems to be lacking. There is no mention of any importance of human involvement in any of the aspects which Computer Science is focusing on.

As the process of creating image representations in CBIR approaches originates more or less directly from different fields of computer science, it is therefore also very influenced by theory and methods from the field of natural sciences. Following from this, much research on CBIR may be regarded as a relatively positivistic approach to science, and as such, concepts like repeatability, reductionism and refutability are fundamental.

Certain aspects of TBIR approaches are also deeply rooted in computer science[15], but when it comes to creating image representations, the logic underlying these approaches are more closely related to information science (and library science) (Yang 2004). The practical implication of this is that the focus of the process is much more on the organisation of both information and information sources so that the end result reflects the value of human involvement, while at the same time serves the needs of users.

Hence, the CBIR approach to the creation of image representations accentuates a positivistic ideal to the process where the goal is an objective interpretation of the image content which is reduced into collections of features. This kind of reductionism stands in contrast to the TBIR approach where factors as perception, knowledge, intuition and experience are seen as valuable assets.

Fjelland (1999) distinguishes between methodological reductionism and ontological reductionism. The former being an uncontroversial and useful scientific strategy, which assumes that the phenomenon is not completely described at the reduced level, while the latter assumes that the observed phenomenon can be completely described at the reduced level (Fjelland 1999). The underlying lines of thought in the TBIR and CBIR approaches to image

[15] Particularly with aspects pertaining to the technological parts: algorithms, query processing, retrieval, evaluation and ranking.

representations may to a certain extent be attributed to each of these forms of reductionism. To my knowledge, researchers in the field of TBIR do not assume that the image content may be completely described using keywords, thus placing the underlying line of thought of the field within the scientific strategy of methodological reductionism. Researchers in the field of CBIR on the other hand, does more often state that image content may be fully described if only some key technological problems (like for instance object recognition) are solved. This line of though does to a larger extent resemble a strategy relying on ontological reductionism.

3.3.1. TBIR and CBIR as Paradigms

As far as viewing the history of image retrieval in a traditional Kuhnian way, the TBIR and CBIR approaches could certainly be said to represent two different paradigms. However, there has been little controversy, and the two fields have co-existed for several years, and have not been competitive fields per se.

As TBIR have a longer tradition than CBIR it is perhaps natural that it has served as a form of benchmark for image retrieval both with regards to what is possible and with regards to potential problems. Following from this, the most noticeable critical remarks have been put forth against traditional TBIR from researchers in the CBIR community.

The large degree of subjectivity and human involvement/manual labour in most TBIR approaches is of many researchers taking the CBIR point of view seen as the most visible problems in image retrieval. The major objections have primarily been associated with the amount of manual labour, and limitations associated with human subjectivity (Faloutsos, Barber et al. 1994; Flickner, Sawhney et al. 1995; Eakins and Graham 1999; Rui, Huang et al. 1999; Brinke, Squire et al. 2006; Smeulders, Worring et al. 2000). The anticipated problem lies in that different people may see and interpret different things when looking at the same object. In fact, the same person may perceive and refer to the same object differently in two different occasions.

In this sense the problem that may occur later is that the idea behind a given formulation may not be understandable unless the circumstances surrounding the situation where the formulation first was used is known (Johannessen 1999:78). The argument against TBIR approaches in this sense is that image representations created and stored in the database in a particular setting in a particular point in time will often not match up with text used as criterion in a query by a user at a later stage. This of course because they do not represent the image contents as such, but rather the annotator’s subjective interpretation of the content.

4. Combining TBIR and CBIR – A Hermeneutic Approach

From the presentation of how image representations are created in each of the two approaches it may seem that the two views represent opposite sides of a coin, thus being mutually exclusive. However in my view, they are not. The CBIR approach to the creation of image representations uses techniques, methods and an underlying line of thought more or less directly originating from the field of Computer Science. The TBIR approach to the creation of image representations on the other hand, uses methods originating form information science and library science. They are thus focusing on different parts of an image, and this could perhaps be used as an advantage in the image retrieval process.

An alternative approach proposed in this essay would be to combine the methods for creating image representations in the CBIR and TBIR approaches in an effort to create representations covering both the perceptual and the conceptual image content. However, this kind of collaborative approaches have been few and far between (Jörgensen 2003:5-6).

4.1. Combining Forces

As computers are constructs, they do not have the ability to instantly "know" what is depicted in an image since they are relying on algorithms, not knowledge, when analysing, transforming and processing images. Following from this, there is thus a mismatch between the way humans and machines "read" images, text and image signatures, and neither the CBIR approach nor the TBIR approach has so far been able to solve this problem effectively on their own. This problem is related to the percept-concept dichotomy briefly discussed above. The CBIR approach is lacking functionality to form a conceptual description of what is "perceived" in the image, while the conceptual description created in the TBIR approach is detached from the actual image content in that it is a product of the interpretation done by an annotator.

A reference to the important link between percept and concept is discussed by Dreyfus et al. (1986). Here, even if actually criticising the notion of Artificial Intelligence, the authors are discussing the ability humans have to more or less gain instant access to relevant memories, knowledge or experiences. The authors see this as some kind of "magic" happening in people's minds based on internal thought processes (e.g. imaginations) or visual impressions (e.g. images or pictures). In effect this may thus be seen as percepts supported by concepts. The main point from the authors is that this is something computers would be unable to replicate.

However, as there is no exactly right memory, it is in fact access to the most appropriate kind of memory, knowledge or experience in a given situation (Dreyfus, Dreyfus et al. 1986). A potential danger when relying on experience and drawing from an available pool of accumulated knowledge is that the conclusion resulting from this interpretation may not necessarily be correct. An example is the observations made by Aztec Indians when the conquistadors were landing in America. The natives described the tall ships as being "clouds" as the sails (being unfamiliar to the eyes of the Indians) bore enough of a resemblance to clouds to be assimilated into that term. This example may perhaps serve as an extreme illustration of the problem of subjectivity.

Seeing this in relation to the link between percepts and concepts, it seems that both humans and computer are prone to errors. This notion of course also applies to the TBIR and CBIR approaches to creating image representations. However, if returning to the Aztec example put in an image retrieval setting, the problem could perhaps be alleviated by CBIR. Providing a computer with an image of sails while calling it clouds does not matter for the CBIR approach as it is focused on low-level features. In essence this means that the computer could return images of sails (if present in the collection) without knowing what it returned, and without regards for what the user actually called the image content.

Concerning the analysis, interpretation and representation of the semantic content of pictures, Hollis (1994) presents some excellent points that can be paraphrased here: There is a contrast between natural signs and conventional symbols, and there is an essential difference between the meaning of words and what people mean by their words (Hollis 1994:161). As can be gathered from the discussion above, the CBIR and TBIR approaches are using natural signs and conventional symbols respectively when creating image representations. A possible approach however, could thus be to combine the two approaches and in that sense make use of *both* natural signs and conventional symbols in the creation of image representations.

Drawing on Latour (1999), the faithfulness of an image representation would in this view be evaluated by its ability to transport the “meaning” of an object through all sorts of transformations. Either it transports it without deformation, and it is deemed accurate, or it transforms it, and it is deemed inaccurate (Latour 1999:248). The transformation process in the two approaches would in this view lead to quite different conclusions. Transformation from picture to image would probably be deemed accurate, while picture to text would probably not. However, as we have seen, the transformation form picture to image is not sufficient to be used for retrieval, and few if any would claim that the collections of low-level features extracted from an image are as accurate as the image itself. In essence, this means that one cannot fully determine how an image looks by using the low-level features or keywords alone because they will never be rich enough or many enough to accurately “recreate” the contents of the image they are representing.

Hence, both the TBIR approach and the CBIR approach are actually facing the same problem. Creating good enough image representations using either approach has proven to be problematic when the representation is to be used by computers to support humans in finding images. However, as we have seen, the two approaches could actually supplement each other in creating an image representation reflecting both the low-level features and the high level concepts.

Combining the two approaches in traditional image retrieval has received some attention from various researchers (Lu 1999; Westerveld 2000; Zhou and Huang 2002; Müller, Ruch et al. 2005), but combining the content-based approach with the text-based approach when creating image representations has to my knowledge not been thoroughly studied even if this could perhaps also help alleviate some of the problems associated with each approach. I have previously tested one approach combining representations consisting of both full-text documents (TBIR) and low-level image features (CBIR) as foundation for image retrieval (Hartvedt 2007), and this study showed very encouraging results.

Seen in relation to the distinction between percept and concept discussed above, both aspects should be covered in the creation of image representations. This could perhaps be, at least in parts, achieved by combining the TBIR and CBIR approaches. Percept would then refer to image data as emphasised by the CBIR approach, whereas concept would refer to the information present in an image representation generated by the TBIR approach. The notion of a symbiotic relationship between percepts and concepts is not new, and Kant stated this quite clearly by saying that concepts without percepts are empty and percepts without concepts are blind (Hollis 1994:89). A viable solution in relation to the creation of image representations would perhaps be to combine CBIR and TBIR and thus create image representations consisting of both “percepts” and concepts. Intuitively this would seem reasonable, but to my knowledge, this approach has not been thoroughly studied.

In their approach to image interpretation presented above, Hilligoss and Howard (2002) assume that it is a human being which actually interpret the pictures and images. However, as we have seen here, some of the tasks on the list, especially those pertaining to recognition of symbols and signs, and the recognition of colours and text, can in fact also very well be carried out by computers.

5. Concluding Remarks

The view taken in this essay is that image analysis and the resulting image representations should be something qualitatively different than quantification of image features and the correlation between them alone. Subjective notions, intuition and cultural factors are also viewed as important in interpreting and describing image content in a meaningful way. This view portrays a very humanistic and hermeneutic science ideal, but also acknowledges the value of “hard” science approaches in that low-level features may aid in describing images in a way not easily done using text.

As Latour (1999) suggests in his book, an important aspect in the process of transforming something into representations actually is to “loose” information (Latour 1999:248), but at the same time gain something from the process. The anticipated positive effect is greater compatibility, standardisation, text, calculation, circulation, and relative universality (Latour 1999:70). This may also be true for image representations created with the use of both low-level features and high-level semantic information.

As we can see from the discussion above, two fundamental differences between humans and computers lie in human’s abilities to form general concepts, and human’s abilities to make use of tacit knowledge and intuition. In this sense problems are bound to occur if we start substituting our minds with computer processors while expecting the same level of performance in tasks requiring cognitive abilities. However, if we are substituting the parts of the process where the performance of computers vastly surpasses our own, the outcome would probably be better.

Concerning the notion of a machine capable of understanding semantics, Dreyfus (1986) suggests this will be an unreachable goal as long as intelligence is understood as abstract reasoning (Dreyfus, Dreyfus et al. 1986). In this sense, putting efforts into improving current methods for creating image representations may prove futile. However, as some computer abilities are far superior to that of humans, for instance identifying and comparing various forms of image textures, a complete understanding of semantics should perhaps not be seen as the ultimate goal in all situations.

The main point being made here is that I believe that the user is still be very important and valuable in providing the system with crucial information on the meaning of various aspects of the depicted contents of a picture. This information will otherwise not be available to the system as the current image representations are not able to capture the information, and even if they were, the machines would probably not be able to understand any of it.

6. Litterature

Baeza-Yates, R. and B. Ribeiro-Neto (1999). Modern information retrieval. New York, ACM Press.

Brinke, W. t., D. M. Squire, et al. (2006). The Meaning of an Image in Content-Based Image Retrieval. CAISE*06 Workshop on Philosophical Foundations on Information Systems Engineering PhiSE '06, Luxemburg, CEUR-WS.org.

Chang, S. K. and A. Hsu (1992). "Image Information Systems: Where Do We Go From Here?" IEEE Transactions on Knowledge and Data Engineering **4**(5).

Colburn, T. (2004). Methodology of Computer Science. The Blackwell Guide to the Philosophy of Computing and Information. L. Floridi, Blackwell Publishing.

Denning, P. J., D. E. Comer, et al. (1989). Computing as a disipline. Report of the ACM Task Force on the Core of Computer Science. New York, The Assosiation for Computing Machinery.

Dervin, B. (1977). "Useful Theory for Librarianship: Communication, Not Information." Drexel Library Quarterly **13**(3): 16-32.

Dorai, C. and S. Venkatesh (2003). "Bridging the semantic gap with computational media aesthetics." Ieee Multimedia **10**(2): 15-17.

Dreyfus, H. L., S. E. Dreyfus, et al. (1986). Mind over machine: the power of human intuition and expertise in the era of the computer, Free Press.

Eakins, J. P. and M. E. Graham (1999). Content Based Image Retrieval: A report to the JISC Technology Applications Program. Newcastle, Inst. for Image Data Research, University of Northumbria.

Eliade, M. (1991). Images and Symbols: Studies in Religious Symbolism. Princeton, NJ, Princeton University Press.

Faloutsos, C., R. , M. Barber, et al. (1994). "Efficient and effective querying by image content." Journal of Intelligent Information Systems **3**(3-4): 231-262.

Fjelland, R. (1999). Innføring i vitenskapsteori. Oslo, Universitetsforlaget.

Flickner, M., H. Sawhney, et al. (1995). "Query by Image and Video Content: the QBIC System." IEEE Computer **28**(9): 23–32.

Gadamer, H. G. (1975). Truth and method London, Sheed & Ward.

Gadamer, H. G. (2003). Forståelsens filosofi. Utvalgte hermeneutiske skrifter. Oslo, Cappelen Akademisk Forlag.

Hartvedt, C. (2007). Utilizing Context in Ranking Results from Distributed CBIR. Norwegian Informatics Conference, NIK, Oslo, Norway.

Heikkinen, H., R. Huttunen, et al. (2000). "And this story is true..." On the Problem of Narrative Truth. European Conference on Educational Research. Edinburgh.

Hilligoss, S. and T. Howard (2002). Visual Communication: A Writer's Guide. New York, Longman.

Hollis, M. (1994). The Philosophy of Social Science: An Introduction. New York, Cambridge University Press.

Jaimes, A. and S. F. Chang (2002). Concepts and Techniques for Indexing Visual Semantics. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and L. D. Bergman. New York, John Wiley & Sons, Inc**:** 497-565.

Johannessen, K. S. (1999). Humanioras vitenskapsfilosofi. Glimt fra vitenskapsfilosofiens hovedområder. K. S. Johannessen. Bergen, Fagbokforlaget.

Jörgensen, C. (2003). Image Retrieval Theory and Research. Laham, Maryland, Scarecrow Press, Inc.

Latour, B. (1999). Pandora's Hope: Essays on the Reality of Science Studies. Cambridge, Massachusetts & London, England, Harvard University Press.

Lew, M. S., N. Sebe, et al. (2006). "Content-based Multimedia Information Retrieval: State of the Art and Challenges." ACM Transactions on Multimedia Computing, Communications, and Applications: 1-19.

Lu, G. (1999). Multimedia database management systems. Boston, Artech House.

Manovich, L. (2001). The Language of New Media. Cambridge, Massachusetts / London, England The MIT Press.

Mitchell, W. (1994). Picture Theory: Essays on Verbal and Visual Representation. Chicago, University of Chicago Press.

Mitchell, W. (1995). Representation. Critical terms for literary study. F. Lentricchia and T. McLaughlin. Chicago., University of Chicago Press**:** 11-22.

Müller, H., P. Ruch, et al. (2005). "Enriching content-based image retrieval with multi-lingual search terms." Swiss Medical Informatics **54**: 6-11.

Neill, S. D. (1987). "The Dilemma of Documentation." The Journal of Documentation **43**(3): 193-209.

Prasad, B. E., A. Gupta, et al. (1987). "A Microcomputer-Based Image Database Management System." IEEE Transactions on Industrial Electronics **IE-34**(1): 83 - 88.

Rui, Y., T. S. Huang, et al. (1999). "Image Retrieval: Current Techniques, Promising Directions And Open Issues." Journal of Visual Communication and Image Representation **10**(4): 39 - 62.

Rui, Y., T. S. Huang, et al. (1998). Relevance feedback techniques in interactive content-based image retrieval. SPIE/IS&T Conf. on Storage and Retrieval for Image and Video Databases San Jose, CA.

Smeulders, A. W. M., M. Worring, et al. (2000). "Content-Based Image Retrieval at the End of the Early Years." IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12): 1349–1380.

Westerveld, T. (2000). Image Retrieval: Content versus Context. RIAO 2000 Conference Proceedings, Paris.

Yang, C. C. (2004). "Content-Based Image Retrieval: A Comparison between Query by Example and Image Browsing Map Approaches." Journal of Information Science **30**(3): 254-267.

Zhou, X. S. and T. S. Huang (2002). "Unifying keywords and visual contents in image retrieval." Ieee Multimedia **9**(2): 23-32.

Østbye, H., K. Helland, et al. (2002). Metodebok for mediefag Bergen, Fagbokforlaget.