

# Difficulties in Image Retrieval

Masashi Inoue<sup>\*</sup>  
National Institute of Informatics  
2-1-2 Hitotsubashi, Chiyoda-ku  
Tokyo, Japan  
m-inoue@nii.ac.jp

## ABSTRACT

The semantic gap is often regarded as a major problem in the field of image retrieval research. In this paper, I will show that there are other important topics that should be addressed for improving the image retrieval utility. Among them, the exploitation of limited information and motivating the use of images are considered to be central to the development of image retrieval.

## 1. INTRODUCTION

### 1.1 Information scarcity

In retrieval, by user studies, it has been revealed that typical user's need for an image is conceptual rather than low-level visual [7]. That means the images are queried by keywords that partially represents the user's search needs. Then, the information is a matchable representation of the image documents against the textual queries; there is the information scarcity unique to the image retrieval, the shortage of textual annotation associated with the images [3]. To alleviate the negative effect of no or few annotations, we should think of the following three approaches: 1) let the users or someone else manually annotate images 2) build a machine that can automatically label images 3) develop a system that works without sufficient annotations.

The first approach is currently practiced. However, as labeling is not interesting for humans and is also time consuming, there needs to be some devices that motivate people to do so. An example of how to motivate users is to turn the image labeling into a game [10]. The second approach entails an unsolvable problem explained in the next section and does not work well in general. The third approach is a compromise that is based on the annotation provided by the first approach and avoids directly addressing the hard problems of the second approach. The third approach seems promising and will be explained in detail with examples. Finally, the paper discusses relatively untouched yet important problem of image retrieval usage expansion.

### 1.2 Semantic gap is not the only problem

The problem of mismatches between the semantics of an image object and its symbolic representation of visual features is termed a semantic gap. In other words, a semantic

gap is the problem that what we see in an image is not what we think of in our mind. This problem resembles the problem known as symbol grounding argued in the context of the advocacy for the combination of the connectionists' approach and the symbolic inferences to derive the semantics from the existing physical entities [2].

The fundamental problem of semantic gap will be recapped as a lack of information within images. Enser et al. named the following example entities that are actually invisible in images, but are of importance to users: time, space, events and significance, abstract and emotive concepts, and unwanted features [1]. Carefully crafted visual features will not help automatically annotating these entities.

### 1.3 Escaping from information scarcity

When first and second approaches do not work well and annotations are insufficient, we should either (a) rely on external resources to compensate for this insufficiency, (b) focus on some domains where visual representations and semantic concepts are directly linked and annotation can be automated, or (c) utilize the visual features of images while trying to avoid the direct linkage between image entities and human conceptualization.

The information insufficiency causes keyword mismatches in which the query words and annotation words cannot be associated through a token-wise comparison. The approach (a) utilizes a pre-build thesaurus or semantic relationship between words extracted from a text corpus that is available outside of target collection. The approach (b), namely the divide-and-conquer approach, leads in the domain specific classification task. Thus far, a successful examples are optical character recognition (OCR) and face recognition, or biometrics that uses finger prints or irises. The approach (c) makes use of the available information as much as possible by exploiting it in different ways. In the following section, I will explain some example scenarios of this approach placing particular emphasis on the utility of visual near-identity or near-duplicate and not the similarity among images.

## 2. EXAMPLE USES OF SUPPLEMENTARY PERSPECTIVE

### 2.1 Insufficiency of annotation words

Images often have only one or two keywords assigned either by their file names or by the tags that bundles the images. One solution to this problem of mismatches between the queries and annotations is the introduction of word associations. When we want to build association models from

<sup>\*</sup><http://research.nii.ac.jp/~m-inoue/index.html>

*Contextual and Social Media Understanding and Usage '08 Dagstuhl, Germany*

a target collection, however, the lack of word co-occurrence due to the lightness in annotations will be a problem in the model estimation. In [6], a visually defined near-identity information between images is used to compensate for the scarcity of text when estimating the word association.

## 2.2 Scarcity of annotation in query language

If the image collection consists of images annotated in multiple languages and among them, few annotations are in the query language, situations where we want to conduct cross-language retrieval might arise. The heterogeneously annotated images will be a harder target than a monolingual collection, because we often have no clue as into which language the query should be translated and how the results should be integrated. That is, the information on the relationship between the annotation languages and the image relevance is limited. For such situations, [5] proposed the use of only the query language and the visual features necessary to access the entire collection of unknown language distributions. The idea is to make nuggets of images before the querying and make use of them when the system has found a relevant image annotated in the query language. In the final ranking, the image that is found will be replaced with the nuggets that contain it and other images annotated in different languages.

## 3. MOTIVATING USERS

### 3.1 Rich communication with images

In contrast to the significant changes in text use after the appearance of printing and digital computers, little change has been observed in the way we use images. Although images have been used to record or appreciate events or personages, and it has become easy to acquire them using cameras, images on arbitrary subjects are still hard to create when compared to texts. Thus, the use of images in modern agile communication should be enhanced by retrieval. As can be seen from examples of emoticons that modify forms of messages, if they can be accessed easily and used in right context, there will be many possibilities that can enhance our ways of communication. However, motivating people who do not search for images is an important yet complicated issue to be defined as a technical research topic.

One approach is to demonstrate when and in what context images are searchable and their use makes a difference. As well as the psychological experiments, information scientific approaches are worth consideration. In the domain of cross-language information retrieval, the use of collaborative frameworks has been suggested, where the creation of successful use scenarios will be cast into the performance comparison amongst systems [4]. Similar approaches might be possible in the domain of image retrieval.

### 3.2 Advanced users and future usages

Another approach to motivate potential users is usage transfer. As we look at the history of technologies, we find that new uses are often developed by professionals who do not hesitate to pay for the new technologies and spend a tremendous time to acquire the ideal results. Therefore, I am going to look at the current professional users of image retrieval to clarify the future direction this field is headed in. Journalists can be considered a good representative of an advanced user of image retrieval. They are keen to use

visuals to attract a reader's attention as it is known that the use of photos will largely influence a reader's behavior [9]. Markkula and Sormunen studied the indexing activities among newspaper journalists [8]. As publishing information has become more universal thanks to the emergence of web publishing, such as blogs, it is likely that ordinary people are more and more involved in the retrieval of images in order to enhance their messages. Photographers, either journalistic or artistic, are another example of advanced users. They take quite a few photos, and then store, select, and show them. Today's ordinary citizens who are not keen to take nice photos just take snaps using their mobiles and send them to a friend with little captions and quickly forget about them. Bridging the gap between the flow of creation and discard of images to the flow of search and use of them is an interesting challenge.

## 4. CONCLUSION

In this paper, I argued that instead of directly tackling the symbol grounding problem, if we consider the actual and potential usages of digital images, we can take the following steps towards making image retrieval practical. First, let people annotate images by means of social and technological tricks. Second, define and focus on the target concepts that can be automatically annotated. Third, make as much use of the available semantic information as possible. I explored the fundamental problem in the third step, namely information scarcity. Then, another important issue of image retrieval, motivating potential users, was discussed.

## 5. REFERENCES

- [1] P. G. Enser, C. J. Sandom, J. S. Hare, and P. H. Lewis. Facing the reality of semantic image retrieval. *Journal of Documentation*, 63(4):465–481, 2007.
- [2] S. Harnad. The symbol grounding problem. *Physica D*, 42:35–346, 1990.
- [3] M. Inoue. On the need for annotation-based image retrieval. In *Proc. Information Retrieval in Context (IRiX)*, pages 44–46, Sheffield, UK, July 2004.
- [4] M. Inoue. The remarkable search topic-finding task to share success stories of cross-language information retrieval. In *New Directions in Multilingual Information Access*, pages 61–64, Seattle, USA, 10 August 2006.
- [5] M. Inoue. Mining visual knowledge for multi-lingual image retrieval. In *DMIR'07*, pages 307–312, Niagara Falls, Ontario, Canada, May 21–23 2007.
- [6] M. Inoue and N. Ueda. Retrieving lightly annotated images using image similarities. In *SAC '05*, pages 1031–1037, March 2005.
- [7] C. Jorgensen. *Image Retrieval: Theory and Research*. Scarecrow Press, 2003.
- [8] M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.
- [9] S. L. Sargent. Image effects on selective exposure to computer-mediated news stories. *Computers in Human Behavior*, 23(1):705–726, January 2007.
- [10] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04*, pages 319–326, 2004.