Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Improving Image Sets Through Sense Disambiguation and Context Ranking

Anthony R. Buck, Walterio W. Mayol
Bristol Robotics Laboratory
Bristol University
Bristol, UK
roja@cs.bris.ac.uk, wmayol@cs.bris.ac.uk

***Abstract*—Current approaches to automatic, class specific, image retrieval from the World Wide Web (WWW) by linguistic query often make use of an image's internal characteristics and file meta-data to augment and improve result accuracy. We propose that, in extension, improvement can be achieved in relevance, noise-reduction and completeness through sense disambiguation and contextual meta-data prepossessing. Our schemes exploits a linguistic ontology identifying query relevant homographs used to construct sense specific keyword sets allowing for enhanced image search and result ranking via the calculation of relatedness between query homographs and image context prior to any additional filtering. Within the paper we investigate different schemes for keyword set construction; ontology exclusive and authority extended, along with three differing ranking mechanisms.**

***Index Terms*—Image retrieval, Meta-data, Vision, Disambiguation, Search**

I. Introduction

The World Wide Web contains a vast corpus of rich media, comment and article which describe a significant part of our world. Unfortunately computational exploitation of this knowledge is restricted due to the unsupervised accumulation of the knowledge which has resulted in a; semantically flat, disorganised, ambiguous and noisy collection of data with no obvious coherent semantic structure. Within this work we look at a subset of this problem, specifically the automated generation of image object class, and more fundamentally semantic sense, specific *prioritized* collections of images (image-sets) from the WWW. The motivation for such better organized image-sets is to enable a range of automated extraction of visual and semantic representations which can be useful to applications in AI, Robotics and HCI to name a few.

Language is often ambiguous, the word "mouse" has numerous discreet meanings, senses, dependant on it's context; a small rodent, a computer input device, a person with a shy disposition, these are all valid definitions for the word.

For a computer program, mapping these different senses is crucial for better automated processing. Take for example the case of a robotic assistant that is tasked with to "find and grasp the mouse".

In order to construct systems which interact more naturally with scene and people, the inclusion of such disambiguation into those systems is critical. This is particularly necessarily the case with a view to image-set collation for use in object class recognition training. If the example of [1] is used and the term "air plane" is entered into a typical image search engine then there is little doubt that despite the returned images (result-set) being noisy, through ranking and classification, a relevant subset of the images can be selected (for [1] by seeding a hierarchical Dirichlet process [2] with the first returned image) which represent typical characteristics of an air plane. However if the user was looking for images of some ambiguous word type, example "plane," the resultant model would, dependant on seeding strategy, either be obviously erroneous due to sense mixing or contain only a single semantic sense of the word. Thus, a more elaborated treatment of the seeding process is needed.

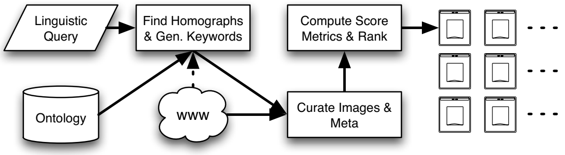


Fig. 1: System Overview Flow. The result is a set of images after the disambiguation of a query term through context ranking.

It is the problem of class ambiguity resulting in single unspecified sense models which are brittle, due to poor seed image selection, which provides our motivation to move from query to sense specific image-set construction enabling sense relevant models and more natural system behaviour.

In this paper we propose SiQIC, Simple Query Image Collation, a prepossessing stage to enable and exploit term sense disambiguation as part of the initial acquisition and as a ranking strategy to provide an ordered multi-set of sense specific images containing labeled positive and negative exemplar for object recognition model training. By utilising a semantic lexical ontology, we are able to disambiguate a term, construct sense relevant search queries and provide a textual model to re-rank the result-set based on the context within which the gathered images are located. This allows us to assign a score signifying the relevance of an image to a semantic sense based on sense of its context rather than simply the inclusion of the ambiguous object class name. We can additionally be discriminating about the senses we enquire as the ontology implicitly provides us with a taxonomic structure for each lemma it contains thus allowing us to identify non-

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

object types with relative ease. An overview of the flow of information in the system is shown in figure 1, and results of ordered image sets for the different rankings considered is shown in figure 2. Note how one of the rankings (Hybrid) presents good examples for a given definition of a query, after both definition and images are automatically gathered by the system.

II. Related Work

Current methodologies for collating model training image data-sets from the WWW fall into three significant strategies:

1) Capturing images from an image search engine, by user query, then filtering the captured result-set fitting each images visual characteristics with those of a discriminating visual model, seeded by a thought good exemplar, and image meta-data.
2) Capturing page results, by user query, using a WWW lexical search engine, identifying and extracting apparent relevant images from pages within the result-set before performing visual characteristic filtration, as per the previous strategy.
3) Capturing images explicitly by visual characteristic filtration, from a large precomputed image set from the WWW, based on an instance or constructed exemplar image of the desired object class instance.

1) The first strategy: Is exemplified by the work of Fergus et al [3], [4], leverages the use of the major image search engines (e.g. Google Images, Picsearch) to capture a potentially noisy set of images (the result-set) roughly related to a specific query term. During this stage heuristics for result-set inclusion are generally unknown thus it must be treated as black box. The result-set is then processed using a pLSA clustering variant, trained using the highest ranked members of the set, to identify a sub-set which adhere, within a threshold, to the characteristics defined within the trained visual model of the class. An analogous approach proposed by L. Fei-Fei et al [1] replaces pLSA with a hierarchical Dirichlet process providing for an iteratively refined filtering model, again seeded by the N highest ranked members of the result-set, capable of allowing more intraclass variance while discriminating ignoring inter-class object instances by continually evolving the filtration model with each image admittance. Again analogous methods were used in [5], [6] though training performed on latest features within the result-set. The post processed result-set typical of these strategies display significant consistent visual characteristics throughout correlating, often strongly, to the initial seed images. This is certainly desirable where a restricted variance result-set of a single object class sense is required, however, there are applications such as object class recognition where it can be useful for a result-set to be composed of considerable intraclass variance including; instance, type, style and pose such that a model capable of recognition under real world conditions of variance can be trained. This is especially true when dealing with classes of naturally high intraclass variance (i.e. chairs, cars, dogs.)

2) The second strategy: Forgoes the use of an image search engine in favour of the more traditional textual search (e.g Google, Yahoo! etc) removing the black box of image ranking evident in the first strategy and instead exposing raw textually term relevant WWW pages as a result-set. This strategy also alleviates the engineering problem of result-set size limitation often imposed by image search engines (Google Images: 1000 results per query for instance.) [7] propose a two phase solution. First a query relevant result-set is acquired with each resultant page harvested for images, over a defined threshold size, and seven context features including; file directory, file name, website title, ten words on either side of the image-link and eleven to fifty words away from the image-link. Secondly, the images are filtered to remove "drawings" from the data-set, under the hypothesis that these are more exposed to noise and significantly more difficult to integrate into a consistent model.

The context meta-data is converted into a binary feature vector such that a feature is true if the associated context contains the search query. The collated images are then ranked based on the associated feature vector. Visual processing then takes place using either pLSA clustering or SVM classification, initialised with the highest ranked result-set images, exploiting SIFT [8] descriptor vectors for the visual comparison. A semi-autonomous solution was proposed by Berg and Forsyth [9] which involved the construction and user-guided partition of term clusters captured from result-set, used to identify pages, and parts there within, which likely contained relevant images which was later evolved by Wang and Forsyth [10] to provide a non-automatic interactive solution which attempts to achieve a better seeding by exploiting what they define as "web resources" explicitly they take a query term and identify a Wikipedia [11] page relevant to that query, explicitly requiring the user to identify the desired sense if ambiguous, they then construct a generative text model from the body of the Wikipedia article. A seed image-set is then selected either from a bank of well tagged example images (Caltech 101, Caltech 256) or from Flickr [12] and used to train a discriminative SVM.

The combination of these two distinct methods of filtration are then applied to result-set derived from a textual search engine and used to sift out text sections relevant to the textual model which contain images acceptable by the SVM with those that pass both criterion accepted into the output result-set. The resultant output of these strategies are much the same as those in the first strategy though with [10] there is an attempt to minimise the arbitrary sense assignment of the model by allowing for interaction from the user to identify a desired sense where available from Wikipedia.

3) The third strategy: Exemplified by [13], [14], [15], is entirely image based and attempts to match representations of image appearance and sketches which are either user selected or constructed. This third strategy is primarily aimed at constructing a different interface for human guided search and does not provide any explicit interface for machine or linguistic use, in fact it can be seen as the second stage in both of the above strategies with the model seeding /

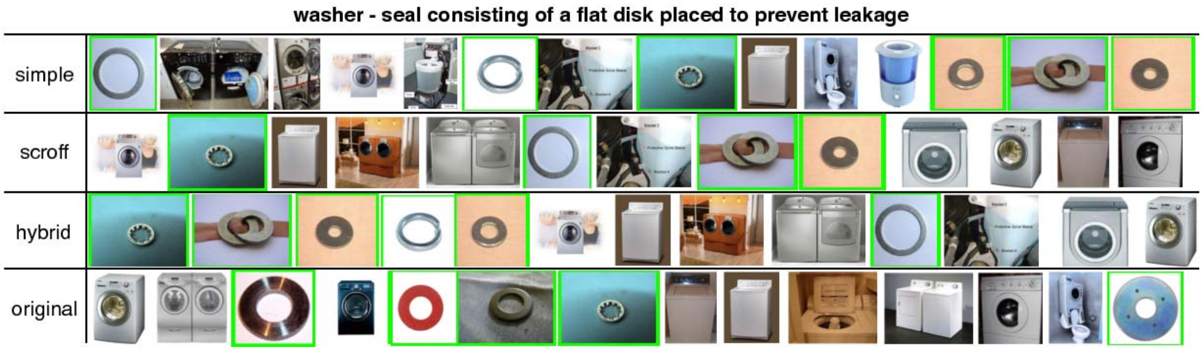

washer - seal consisting of a flat disk placed to prevent leakage
simple
scroff
hybrid
original

Fig. 2: A comparison of rankings, for an image set captured using simply query word, through the use of differing scoring metrics set to score for the definition “seal consisting of a flat disk placed to prevent leakage.”

bootstrapping stage entirely removed and replaced with a user content selection / specification for seeding the classification model manually. Although this may provide advantageous from a HCI perspective or for specialised tasks it is not a relevant strategy for the task outlined within this paper.

III. Approach

Algorithm 1 Overview: Disambiguation and Ranking

```
Acquire simple query term.
Locate query term relevant noun homographs of type; animal, artifact, food, object, plant or substance.
for all homographs do
    Construct sense keyword sets.
    Develop a query string and perform an image search using it.
    for all results do
        Generate image context features.
        Calculate a context-homograph relatedness score.
    end for
    return image-set prioritised by score
end for
```

SiQIC acts as a prepossessing framework, a combination of two distinct stages; query sense disambiguation and result-set ranking. Alg.1 and Fig.1 illustrate how these two phases are combined to produce a ranked result-set which can then be passed onto a visual filtration system in an overview of the complete process.

A. Term Sense Disambiguation

Given a simple *query term*, associated with the name of some object class, we extract the term from WordNet [16] and identify homographs (either homonyms or heteronyms) which are of noun form. In addition, we also cull those which don’t have semantic lexical type of; animal, artifact, food, object, plant or substance, this allows us to specifically identify those homographs which are relevant for our task and discard the rest.

B. Sense Keyword Extraction

There are two differing schemes towards sense keyword extraction within the proposed system; ontology exclusive and authority extended. Each of these are explored further within the experimentation stage (See.IV.)

1) Ontology Exclusive: For each homograph we exploit semantic knowledge from the ontology, we construct two clean (See. III-B3) sense specific keyword sets; *lemmaSet* and *senseSet.* The *lemmaSet* consists of the lemma, word, of the current homograph along with each term identified as being either a hypernym or hyponym of the homograph. A “Gloss” is similar to the dictionary definition of a specific word, often including exemplar sentences of word use. The *senseSet* contains all words found within the glosses of terms identified as being either a hypernym or hyponym of the homograph along with those directly from the homographs gloss.

2) Authority Extended: In addition to the sense specific keyword sets; *lemmaSet* and *senseSet* developed within the ontology exclusive scheme the authority extended scheme attempts to extend the quantity of sense specific keywords available. By seeking out sense relevant entries within authority websites; Wikipedia, Encarta and Britannica, we construct an additional keyword set *authSet.* For each authority website a host restricted search query is constructed consisting of a disjunction of the *lemmaSet* as title requirements and a disjunction of the *senseSet* as desirable words. The query is used with Google to identify the entry most relevant within the authority site. The core content of the entry is extracted, formatting and site specific information can be removed due to site known convention, and appended to the *authSet.* Once fully constructed the *authSet* is cleaned as with the other keyword sets.

3) Keyword Set Cleaning: Keyword cleaning is the process undertaken to sanitise each keyword list, at each step of the process Porter stemming [17] is used to return the keyword to a comparable root form. First any stop-words are removed from the keyword list, we identify as stop-words the most frequent 571 words in the English language [18] along with the lemma of the current homograph. The keywords are then ordered by frequency (highest first) and any intra-list (more than one instance of the same keyword in the list) and inter-list (any keywords found in the same keyword list of other homographs of the *query term*) duplicates are removed. The resultant is a stemmed and stop-worded keyword set with no inter-set no duplication.

Although we use English here, a similar procedure can be formulated for other languages.

C. Image Search String Generation

An image search string is then generated for each homograph consisting of the *lemmaSet*, the homographs semantic lexical type (if not "artifact" or "object".) If the query string only contains a single keyword then keywords which appear in both *glossSet* and *lemmaSet* are appended and if even then the query string only contains a single keyword and authority extended keywords are being used then keywords which appear in both *glossSet* or *lemmaSet* and *authSet* are appended. It is this query string which is used as an extended query into any number of online image search engines allowing for the construction of an image-set. The resultant is a set of image-sets, ideally, each specific the initial query term.

D. Context Feature Extraction

For each image collated we extract eight features relevant to the images context consisting of the 7 features proposed in [7]: *fileDir*, *fileName*, *websiteTitle*, *imageTitle*, *imageAlt*, *context10*, *contextR* and the additional *contextA*.

fileDir contains the path, minus the host and the file name, to the image, *fileName* is self explanatory, *websiteTitle* contains the contents of the main "title" tag of the context page, *imageTitle* and *imageAlt* refer to the "title" and "alt" tags of the "img" tag associated with the image from the context page. The context features; *context10*, *contextR* contain the initial ten and fifty pre and post "img" tag words after cleaning (See. III-B3) and *contextA* contains a clean set of all words on the context page including those found within the "meta" tag structure. A system of parsing is used to ensure all HTML content and other WWW specific content is decoded and removed where appropriate. As an additional filter any images which fail to appear within there context, i.e. content which has changed since search indexing, or contexts which contained less then 20 words post cleaning were removed from the image-set as to not expose the system to dated or low detail content.

E. Ranking

At this stage we have a homograph specific image-sets with images ranked in the order specified by the image search engine, *original ranking*. However in order to exploit our additional information further it is possible to re-rank the images based on a context-homograph relatedness score. We have three different scoring strategies used for experimentation; *Simple* (See. III-E1), *Schroff-like* (See. III-E2) and *Hybrid* (See. III-E3.) Each scheme makes use of both the homograph and context specific keyword sets to re-rank the images linguistically, identifying high likelihood images and promoting them.

1) Simple: The simple score metric (See. Alg.2) counts the instances of each keyword within the homographs *senseSet*, and optionally *authSet* if constructed, that is identified within the image's *contextA*.

Algorithm 2 Simple Scoring Strategy

```
score = 0
for all keyword in senseSet and optionally authSet do
    if contextA contains keyword then
        increment score
    end if
end for
return score
```

2) Schroff-like: Schroff et al proposed the use of a seven dimension binary *sense vector* where each dimension is attributed to one of the seven image context features; *fileDir*, *fileName*, *websiteTitle*, *imageTitle*, *imageAlt*, *context10*, *contextR*. Each feature is checked for *query word* inclusion and where identified the particular vector is marked as true. Schroff-like scoring similar to that proposed, incrementally scores based on the inclusion of the homograph lemma within each of the image context features, utilising stemming is to allow for fairer string comparison. This results in a integer score 0..7 based on the number of matches, instead of a *sense vector.*

Algorithm 3 Schroff-like Scoring Strategy

```
score = 0
featureArray = [fileDir, fileName, websiteTitle, imageTitle,
imageAlt, context10, contextR]
for all feature in featureArray do
    if feature contains lemma then
        increment score
    end if
end for
return score
```

3) Hybrid: Hybrid scoring allows for the additional sense information; *senseSet* and optionally *authSet* if constructed, to be exposed to the Schroff-like scoring strategy. This is done by allowing each keyword within *senseSet* and *authSet*, where applicable, along with homograph lemma to be checked for inclusion within each of the image context features, again utilising stemming is to allow for fairer string comparison. This allows for the exploitation of the additional sense specific keywords in identifying sense-relevance while placing a lower-bound on the score such that even if the two additional keyword sets contain minimal additional keywords the strategy will return the Schroff-like score.

IV. Experiments

In order to assess the effectiveness of the proposed system, thirteen common household objects with ambiguous names were selected: basket, bat, bolt, bulb, disc, file, mouse, plane, recorder, shoe, spade, tank, washer. Each object was entered into the system independently through two complete runs, the

Algorithm 4 Hybrid Scoring Strategy

```
score = 0
featureArray = [fileDir, fileName, websiteTitle, imageTitle,
imageAlt, context10, contextR]
for all feature in featureArray do
  for all keyword in senseSet and optionally authSet do
    if feature contains keyword then
      increment score
    end if
  end for
end for
return score
```

first with ontology exclusive and the second with keyword extended keyword generation. For every image captured each of the three scoring metrics was used allowing for comparison. Due to the usage policies and restrictions of the image search engine API, and the systems removal of outdated and poor context, 35-50 images were returned per homograph. A detailed evaluation of the resultant image-sets can be found in the results section. In addition to the overall results however a number of more specific hypothesis were proposed and tested against the output.

A. Hypothesis: Sense specific querying improves object data-set completeness

It is possible to test this hypothesis by identifying the number of distinct senses, those identified as homographs, which are represented in a simple control set of images returned by the image search engine when given the object name as the query term. This was performed for each of the selected object names then compared against the number of disambiguated image-sets which provide at least a 50% content of sense representative images, both ontology exclusive and authority extended.

1) Results: The results (Tbl.I) show with an average of 2.92 definitions per object name the average coverage of senses within the control set is 55.26% where as 78.85%/75.00% of the senses are covered by the disambiguated sets with a requirement of over 50% consistency and 83.97%/83.33% if the requirement is reduced to 30% consistency.

2) Discussion: The disambiguated results, for the tested object list, shows that sense coverage can certainly be improved by altering the querying strategy to look for specific sense images. This is predictable due to the increase in information provided to the search engine itself through the inclusion of additional sense keywords. This however is countered by the reduction when moving from ontology exclusive to authority extended. This reduction is indicative of the terms appended to the search by the authority keyword set being of a higher generality or less relevancy and thus diluting the other query terms allowing more irrelevant images to be returned effecting the quantity of consistent images to below the threshold 50%. The results at 30% requirement back this up with the values converging.

The increase of sense consistency however is not the only result, it is obvious that due to the segmentation of the disambiguated image-sets the consistent results (excess of 50% of the set) are automatically labeled with their specific sense unlike the control set which, though may cover multiple sets, has no sense differentiation within it's contained images. This labeling can help identify inconsistent images by providing true-negative exemplar from the other image-sets conceivably improving the training of a visual classifier.

B. Hypothesis: Sense specific querying can reduce intra-set overlap through term disambiguation

Similarly it is possible to test that the system can reduce intra-set overlap through term disambiguation by comparing the quantity of distinct senses which are represented in a simple control set of images with the quantity of senses identified within disambiguated image-sets when using both both ontology exclusive and authority extended keyword generation.

1) Results: The results (Tbl.II) show that on average 61.54% of the image-sets, captured using the query term, alone contained sense overlap within them. In comparison if sense specific queries were constructed via term disambiguation only 29.9%/26.28 of the returned image sets contained any sense overlap.

It is also possible to see that specific homographs can heavily effect the average sense overlap within a disambiguated set, example; shoe and bulb. These particular sets, notably bad within the ontology exclusive set, are effected by the vast prominence of a singular meaning over other meanings, particular in this case that of the footwear for the term shoe and the electric light for the term bulb. This prominence along with the large variance within the object described contributes to increased likelihood of set overlap.

TABLE II: Sense Overlap Result

			w/o. Auth.	**w. Auth.**
Object Name	**Senses**	**Cont. Olap?**	**Dis. Olap?**	**Dis. Olap?**
basket	2	0.00%	0.00%	0.00%
bat	4	0.00%	25.00%	25.00%
bolt	4	100.00%	50.00%	0.00%
bulb	3	0.00%	66.67%	66.67%
disc	3	100.00%	33.33%	33.33%
file	2	0.00%	0.00%	0.00%
mouse	2	100.00%	50.00%	50.00%
plane	3	0.00%	33.33%	66.67%
recorder	2	100.00%	0.00%	0.00%
shoe	4	100.00%	75.00%	50.00%
spade	3	100.00%	0.00%	0.00%
tank	4	100.00%	0.00%	0.00%
washer	2	100.00%	50.00%	50.00%
Mean.	2.92	61.54%	29.49%	26.28%

2) Discussion: It is sensible to predict, as with the previous hypothesis that improvement would be made due to the increase in information available to the search engine used. An improvement in excess of 100% identifies that, though some intra-set ambiguity still remains, it should be easier to identify with an incremental learning visual classifier due to

TABLE I: Sense Completeness Results

				w/o. Auth.		w. Auth.	
Object Name	**Senses**	**Cont.**	**Prec.**	**Dis. (50%)**	**Prec.**	**Dis. (50%)**	**Prec.**
basket	2	1	50.00%	2	100.00%	2	100.00%
bat	4	1	25.00%	2	50.00%	3	75.00%
bolt	4	2	50.00%	3	75.00%	2	50.00%
bulb	3	1	33.33%	3	100.00%	3	100.00%
disc	3	2	66.67%	1	33.33%	1	33.33%
file	2	1	50.00%	2	100.00%	1	50.00%
mouse	2	2	100.00%	2	100.00%	2	100.00%
plane	3	1	33.33%	3	100.00%	3	100.00%
recorder	2	2	100.00%	2	100.00%	2	100.00%
shoe	4	2	50.00%	3	75.00%	3	75.00%
spade	3	2	66.67%	2	66.67%	2	66.67%
tank	4	2	50.00%	3	75.00%	3	75.00%
washer	2	2	100.00%	1	50.00%	1	50.00%
Mean.	2.92	1.62	59.62%	2.23	78.85%	2.15	75.00%

increased positive exemplar along with the provided sense specific negative exemplar typically depicting those images which are causing the set overlap. The improvement when using authority keywords can also be explained as although the added keywords may increase the probability of non-valid images entering a set (Sec.IV-A,) due to there relatively reduced sense relevance, they may still provide additional inter-sense discrimination information to the search engine. The object term "bat" provides a good example, as it is much improved between the two sets, and the definition "a club used for hitting a ball in various games" where the authority pages identified "baseball" as a frequent keyword, this may result in additional invalid images, i.e. baseball cards, but may further the general resultant set from other senses, i.e. "a flying mammal."

C. Hypothesis: Context sense ranking increases the opportunity of selecting a sense relevant prime image instance

One of the main objectives of this paper is ensuring we can reliably identify a single image (or a reduced set) which can act as a seed image for further vision-based discriminative filtering, a prime image instance. To identify to what extent this has been achieved for each homograph, in each selected object name, we look at the rate at which a sense valid image appears ranked first for each of the scoring schemes for both ontology exclusive and authority extended keyword generation.

1) Results: The result-set (Tbl.III) visualises the first result, prime image instance, consistency precision when linguistic context ranking is applied to the returned image-sets from both keyword generation schemes using each of the three scoring metrics; Simple, Schroff-like, Hybrid, along with the ranking given by the search engine itself, Original. Simple scoring fairs the worst at 69.23%/57.05% followed by Original 71.15%/63.46% then Shroff-like at 71.15%/73.08% with the best results being produced by the Hybrid metric with 85.26%/75.64%. The discrepancy between the two Shroff-like percentages is attributable to the differing image sets constructed and not anything inherently different with the scoring due to the scheme only accounting for the query term and thus not directly being effected by the additional *authSet* keywords.

2) Discussion: The results provide encouraging evidence that a gain in consistent selection precision can indeed be made by performing re-ranking, however it is important to ensure that the correct scoring metric is used.

Simple scoring (III-E1), predictably fared the worst, resulting in substantially worse precision than the original rank applied by the search engine.

Simple scoring only concerns itself with whether a keyword exists within a whole image context page, it's locality within that page, especially relative to the image in question, is unaccounted. This leads directly to poor scoring accuracy when provided with context pages which may well be sense specific but which contain numerous images some of which may be non sense specific, causing potential escalation of those images.

Schroff-like ranking (III-E2) fared better producing a similar precision to that of the Original rank but being more stable when the *authSet* keywords are introduced. This is to be expected as the Schroff-like metric is ignorant of the additional *authSet* keywords however the search query used by the search providers ranking will be exposed to those additional keywords.

The Hybrid metric (III-E3) however improved the precision of selection by 12-15% over those of the Original ranking identifying that by re-ranking using additional sense specific information when applied to specific localities within an images context can improve the likelihood of selecting a sense consistent prime instance image.

As seen before however the addition of authority keywords can heavily effect the result. By allowing the additional keywords in authSet the three metrics which made use of those keywords typically dropped precision by 7-10%. As for previous experiments this is conceivably due to the increase of less sense specific keywords from which metric comparison is made resulting in scores which reflect more the similarity of pages to those authority pages than the similarity of there contained senses.

TABLE III: Prime Instance Results

		w/o. Auth.				w. Auth.			
Object Name	**Senses**	**Simple**	**Schroff**	**Hybrid**	**Original**	**Simple**	**Schroff**	**Hybrid**	**Original**
basket	2	0.00%	50.00%	100.00%	100.00%	0.00%	50.00%	50.00%	100.00%
bat	4	50.00%	75.00%	75.00%	75.00%	75.00%	75.00%	25.00%	50.00%
bolt	4	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
bulb	3	33.33%	66.67%	66.67%	66.67%	33.33%	66.67%	66.67%	66.67%
disc	3	100.00%	33.33%	100.00%	33.33%	66.67%	33.33%	66.67%	33.33%
file	2	100.00%	100.00%	100.00%	100.00%	50.00%	100.00%	100.00%	50.00%
mouse	2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
plane	3	66.67%	100.00%	66.67%	100.00%	66.67%	100.00%	100.00%	100.00%
recorder	2	50.00%	50.00%	100.00%	100.00%	0.00%	100.00%	100.00%	100.00%
shoe	4	75.00%	75.00%	75.00%	50.00%	50.00%	50.00%	50.00%	25.00%
spade	3	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
tank	4	75.00%	75.00%	75.00%	50.00%	50.00%	75.00%	75.00%	50.00%
washer	2	100.00%	50.00%	100.00%	0.00%	100.00%	50.00%	100.00%	0.00%
Mean.	2.92	69.23%	71.15%	85.26%	71.15%	57.05%	73.08%	75.64%	63.46%

V. Results

In this section the overall results produced by combinations of ontology exclusive and authority extended keyword generation along with each of the three scoring metrics are evaluated. In addition the difference on the identification and removal of sets with zero consistency can have on the overall results is reviewed. Within table IV and figure 3a the mean percentile sense consistency for each metric and keyword scheme is presented firstly displaying those results based on the full range of each homograph for each of the selected object names then continuing on display a second set of results, w/o. ZC. This second set of results illustrates the same system runs, however, selected object name homographs which return an image-set with zero consistency, no image displayed is indicative of the sense of the homograph, are removed prior to analysis. Thus the results are predictive of performance obtainable via removal, post visual filtering, of image-sets containing significantly fewer visually consistent images than other sense specific image-sets from the same object name.

Within both sets of results, similar performance is obtained, regarding scoring metric, as within the identification of prime instances. In addition, given that the images-set size is restricted to the first *n* ranked images, as *n* increases all of the precisions tend to converge, tending towards an average of 60%, this is consistent with the true average consistency of all object sense image-sets of 61.05%/61.13% (w. Auth). The Hybrid metric tends to converge at the lowest rate and provides, given the result-set of between 35 and 50 images after system validation, a performance increase for sets of up to 15 images of 13.51-7.39%/10.8/4.23%. Lesser results can be seen for the Schroff metric and a loss is typically seen for the Simple metric when compared with the Original ranking. This is consistent with trends identified within the other experiments. Reduction is again seen in performance when authority keywords are used by the metrics this is predictable due to the same reduction in keyword specificity identified when selecting prime images resulting in the escalation of images with increased authority page similarity not necessarily sense relevance.

The results post removal of image-sets with zero consistency show a marked improvement of an average of 7.95%/7.35% with the Hybrid (w/o. Auth) gaining an additional 9.83% over it's ZC inclusive prior. There is also significant alteration to the point of conversion in-line with the average increase. This indicates that the identification of these sets would provide useful further improvement.

VI. Conclusion

In this paper three metrics for textual context scoring were introduced along with two schemes for additional sense specific keyword generation and protocol to perform sense enhanced searching. The results show that given the optimal combination of these parts, ontology-exclusive keywords and Hybrid metric, sense relevant prime-instance identification rate can be increased to excess of 81% and partial image-sets can be constructed with significant consistency improvements over those simply provided by search engines.

By applying sense relevant query extension to image searching reduction in inter-set noise and overlap can also be observed. This is made possible via the automated generation of sense specific keywords through the use of an ontology along with it's optional extension through WWW resource.

Thus the systematic improvements provide evidence for the inclusion of disambiguation as part of any system which intends to exploit WWW retrieved image sets requested through natural ambiguous language.

Acknowledgment

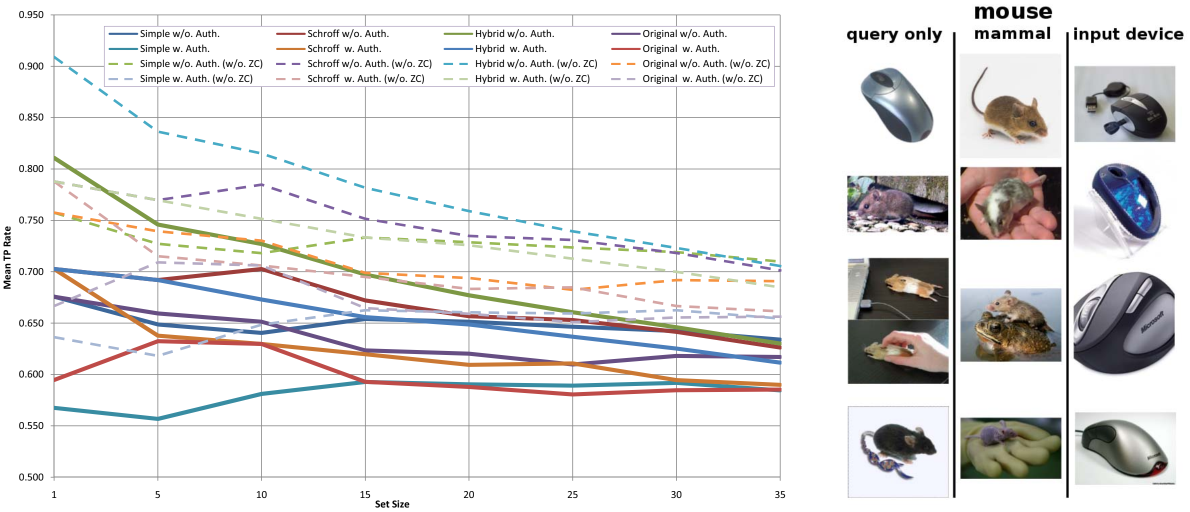
The authors would like to thank the EPSRC for there supportive funding of the first authors DTA grant.

References

[1] L. Li, G. Wang, and L. Fei-Fei, "Optimol: automatic online picture collection via incremental model learning," in *Proc. Computer Vision and Pattern Recognition*, vol. 2, 2007.

[2] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.

[3] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from google's image search," *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2, pp. 1816–1823 Vol. 2, Oct. 2005.



(a) Image-set Consistency Results.

(b) Original v. Disambiguated Set Example

Fig. 3: Image-set Consistency Results: Dashed lines represent those sets which have had zero consistency homographs removed. Original v. Disambiguated Set Example: Query only contains the first 4 items returned by simply the query "mouse", the further two sets are the disambiguated sense specific results after hybrid metric ranking using ontology-exclusive keyword generation.

TABLE IV: Image-set Consistency Results

			1	**5**	**10**	**15**	**20**	**25**	**30**	**35**
	w/o. Auth.	**Simple**	67.57%	64.86%	64.05%	65.41%	65.14%	64.65%	64.23%	63.40%
		Schroff	70.27%	69.19%	70.27%	67.21%	65.68%	65.30%	64.14%	62.63%
		Hybrid	81.08%	74.59%	72.70%	69.73%	67.70%	66.05%	64.59%	63.01%
		Original	67.57%	65.95%	65.14%	62.34%	62.03%	60.97%	61.80%	61.70%
	w. Auth.	**Simple**	56.76%	55.68%	58.11%	59.28%	59.05%	58.92%	59.19%	58.46%
		Schroff	70.27%	63.78%	62.97%	61.98%	60.95%	61.08%	59.46%	59.00%
		Hybrid	70.27%	69.19%	67.30%	65.59%	64.86%	63.68%	62.52%	61.16%
		Original	59.46%	63.24%	62.97%	59.28%	58.78%	58.05%	58.47%	58.53%
w/o. ZC	**w/o. Auth.**	**Simple**	75.76%	72.73%	71.82%	73.33%	72.88%	72.36%	71.92%	71.00%
		Schroff	78.79%	76.97%	78.48%	75.15%	73.48%	73.09%	71.82%	70.13%
		Hybrid	90.91%	83.64%	81.52%	78.18%	75.91%	73.94%	72.32%	70.56%
		Original	75.76%	73.94%	73.03%	69.90%	69.39%	68.24%	69.19%	69.09%
	w. Auth.	**Simple**	63.64%	61.82%	64.85%	66.26%	66.06%	65.94%	66.26%	65.45%
		Schroff	78.79%	71.52%	70.61%	69.49%	68.33%	68.48%	66.67%	66.15%
		Hybrid	78.79%	76.97%	75.15%	73.33%	72.58%	71.27%	70.00%	68.48%
		Original	66.67%	70.91%	70.61%	66.46%	65.91%	65.09%	65.56%	65.63%

[4] R. Fergus, P. Perona, A. Zisserman, and D. E. Science, "A visual category filter for google images," in *In Proc. ECCV*, 2004, pp. 242–256.

[5] N. Ben Haim, B. Babenko, and S. Belongie, "Improving web-based image search via content based clustering," 2006, p. 106.

[6] H. Zitouni, S. Sevil, D. Ozkan, and P. Duygulu, "Re-ranking of web image search results using a graph algorithm," *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, Dec. 2008.

[7] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1–8.

[8] D. G. Lowe, "Object recognition from local scale-invariant features," 1999, pp. 1150–1157.

[9] T. Berg and D. Forsyth, "Animals on the web," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006.

[10] G. Wang and D. Forsyth, "Object image retrieval by exploiting online knowledge resources," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[11] "Wikipedia, the free encyclopedia." [Online]. Available: http://www.wikipedia.org

[12] "Flickr - photo sharing." [Online]. Available: www.flickr.com

[13] Y. Chan, Z. Lei, D. Lopresti, and S. Kung, "A feature-based approach for image retrieval by sketch."

[14] G. Ohashi and Y. Shimodaira, "Query-by-sketch image retrieval using relevance feedback," vol. 6051, p. 60510Z, 2005.

[15] A. Grigorova, F. De Natale, C. Dagli, and T. Huang, "Content-based image retrieval by feature adaptation and relevance feedback," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 9, no. 6, p. 1183, 2007.

[16] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, May 1998. [Online]. Available: http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&amp;path=ASIN/026206197X

[17] C. Van Rijsbergen, S. Robertson, M. Porter, B. L. Research, D. Dept, and U. of Cambridge. Computer Laboratory, *New models in probabilistic information retrieval*. Computer Laboratory, University of Cambridge, 1980.

[18] J. Savoy, "The english language stopword list." [Online]. Available: http://members.unine.ch/jacques.savoy/clef/