PicAlert!: A System for Privacy-Aware Image Classification and Retrieval

Sergej Zerr*, Stefan Siersdorfer*, Jonathon Hare**
*L3S Research Center, Hannover, Germany
{zerr,siersdorfer}@L3S.de
**Electronics and Computer Science, University of Southampton, Southampton, UK
jsh2@ecs.soton.ac.uk

ABSTRACT

Photo publishing in Social Networks and other Web2.0 applications has become very popular due to the pervasive availability of cheap digital cameras, powerful batch upload tools and a huge amount of storage space. A portion of uploaded images are of a highly sensitive nature, disclosing many details of the users' private life. We have developed a web service which can detect private images within a user's photo stream and provide support in making privacy decisions in the sharing context. In addition, we present a privacy-oriented image search application which automatically identifies potentially sensitive images in the result set and separates them from the remaining pictures.

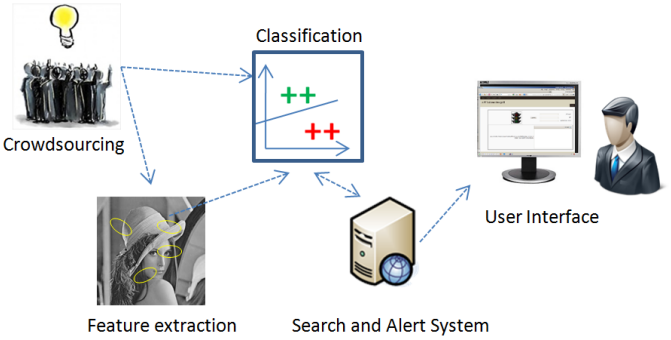

Figure 1: System architecture overview.

1. INTRODUCTION

With increasing availability of content sharing environments such as Flickr, and YouTube, the volume of private multimedia resources publicly available on the Web has drastically increased. In particular young users often share private images about themselves, their friends and classmates without being aware of the consequences such footage may have for their future lives [1, 5]. Users of photo sharing sites often lack awareness of privacy issues. Our recent study [6] revealed that up to 20% of publicly shared photos on Flickr are of sensitive nature. Existing sharing platforms often employ rather lax default privacy configurations, and require users to manually decide on privacy settings for each single resource. Given the amount of shared information, this process can be tedious and error-prone. This is especially true for large batch photos uploads. Furthermore, image search engines do not provide the possibility to directly search for private images which might already be available on the web.

In this work we demonstrate the PicAlert! privacy-oriented image search application. PicAlert! is able to identify and isolate images in a Flickr result set that are potentially sensitive with respect to user privacy. The application is based on a web service that automatically identifies a privacy degree of an image through classification of the content and context of the image. It could be directly integrated into social photo sharing applications like Flickr or Facebook, or into browser plugins in order to support users in making adequate privacy decisions in image sharing. Thus, the application illustrated in the demo is two-fold: warning the user about uploading potentially sensitive content on the one hand (cf. Figure 2) and privacy-oriented search on the other hand (cf. Figure 3).

We are aware that building alarm systems for private content and enabling privacy-oriented search can be seen as contradicting goals; privacy-oriented search is not negative per-se, as it can be used for retrieving private content users are comfortable to share, and, more importantly, can help with the early discovery of privacy breaches. However, as with almost every technology, it requires sensible handling and constructive usage.

2. SYSTEM ARCHITECTURE

This section describes the main components of the PicAlert! system. The system architecture is illustrated in Figure 1. Firstly, through *crowd-sourcing*, we build a training set of private and public images. In the next step we extract visual, and, if available, textual *features* which provide hints for the privacy degree of an image. We then train a SVM *classifier* which is used by our Search and Alert system for identifying potentially sensitive visual content. Finally, the user can access the application from arbitrary *clients* including desktops and mobile devices. In the following we provide a brief overview of the system components and show how results are presented to the user. A fully detailed description of the underlying scientific approach can be found in our recent work [6].

CIKM '12 Maui, Hawaii USA

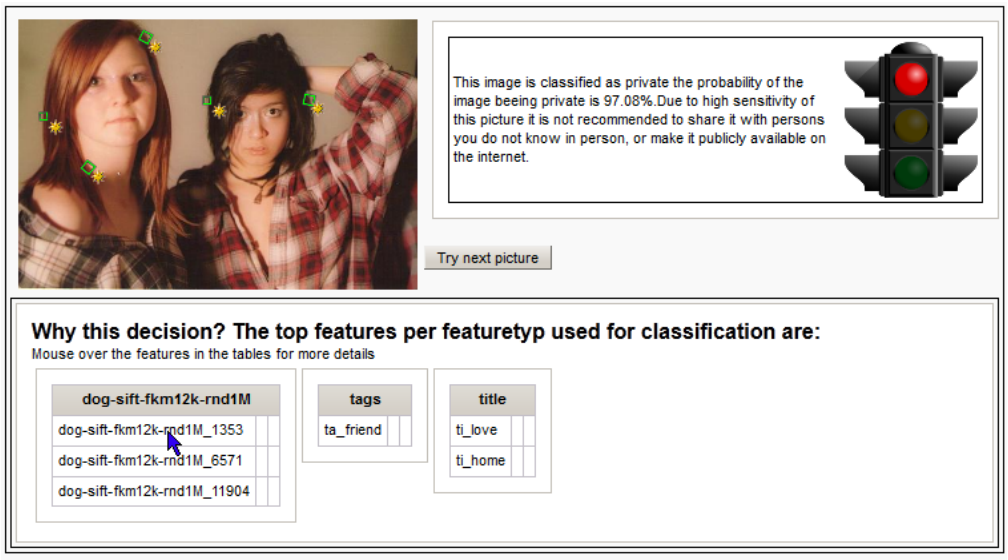

Figure 2: Web service GUI for privacy-oriented image classification.

2.1 Data & Crowdsourcing

In order to obtain an appropriate dataset with labeled private and public image examples, we performed a user study in which we asked external assessors to judge the privacy of photos available online. To this end, we crawled 90,000 images from Flickr, using the "most recently uploaded" option to gather photos uploaded in a time period of 4 months. As labelling a large-scale image dataset requires considerable manual effort, we conducted the user study as an annotation game. At each step of the game we presented five photos to a participant of the study. For each photo, the participants had to decide if, in their opinion, the photos belonged to the private sphere of the photographer. Specifically, we asked the participants to imagine that images presented to them were photos they took with their own cameras, and mark these images as "private", "public", or "undecidable". We provided the following guidance for selecting the label: "*Private* are photos which have to do with the private sphere (like self portraits, family, friends, your home) or contain objects that you would not share with the entire world (like a private email). The rest are *public*. In case no decision can be made, the picture should be marked as *undecidable*." Over the course of the experiment, 81 users between ten and 59 years of age labeled 37,535 images.

We examined the user choices and defined a labelling threshold. Each picture was labeled private or public if at least 75% of the judges were of the same opinion. Overall the dataset contained 4,701 images labeled as private, and 27,405 images labeled as public; the remainder were marked as undecidable.

2.2 Features

Digital images are internally represented as two-dimensional arrays of color pixels. This representation is difficult to use directly for classification because it is highly multidimensional and subject to noise. Instead, a process known as feature extraction is typically used to make measurements about the image content. Image features come in many forms, from the very low-level, to so-called high-level features. Low-level features are typically formed from statistical descriptions of pixels, whilst high-level features are those that have a directly attributable semantic meaning. For this application, we have selected a range of image features that could potentially be used in building a classifier that can discriminate public and private images automatically. In particular we observed that the occurrence of *faces* in a picture is strongly associated with a high degree of privacy although a considerable number of faces also can be found in public images. Intuitively, *color* may be an indicator for certain types of public and private image. For example, public images of landscape scenes are very likely to have a specific color distribution. The *edges* within an image are a very powerful feature for discriminating between different in/outdoor types of scene, and are useful for privacy classification. Finally the *SIFT* descriptor [4] turned out to be the most powerful feature for our application. Private and public photos typically tend to be taken in specific contexts. For example, pictures can be taken in public places like stadiums, supermarkets and airports, or in private places like home, car, or garden. Accordingly the object parts contained in a photo, like sport equipment, furniture, human and animal body parts are represented as SIFT features and could be different and thus give us insights about an image's privacy. For efficiency reasons we limited the visual features used for the demonstration application to face detection and SIFT features. Additionally, we made use of textual features including the image tags and title.

2.3 Classification

We obtained a balanced training set by randomly restricting the initial image set described in Section 2.1 to a subset of 9402 images with an equal number of public and private images. The balanced set helps to capture general classifier properties independently of the a-priori class probabilities of the dataset. In the next step we built classifiers using the SVMlight [2] classification software. The results of the classification experiments for selected visual features described in Section 2.2 are presented in Section 3

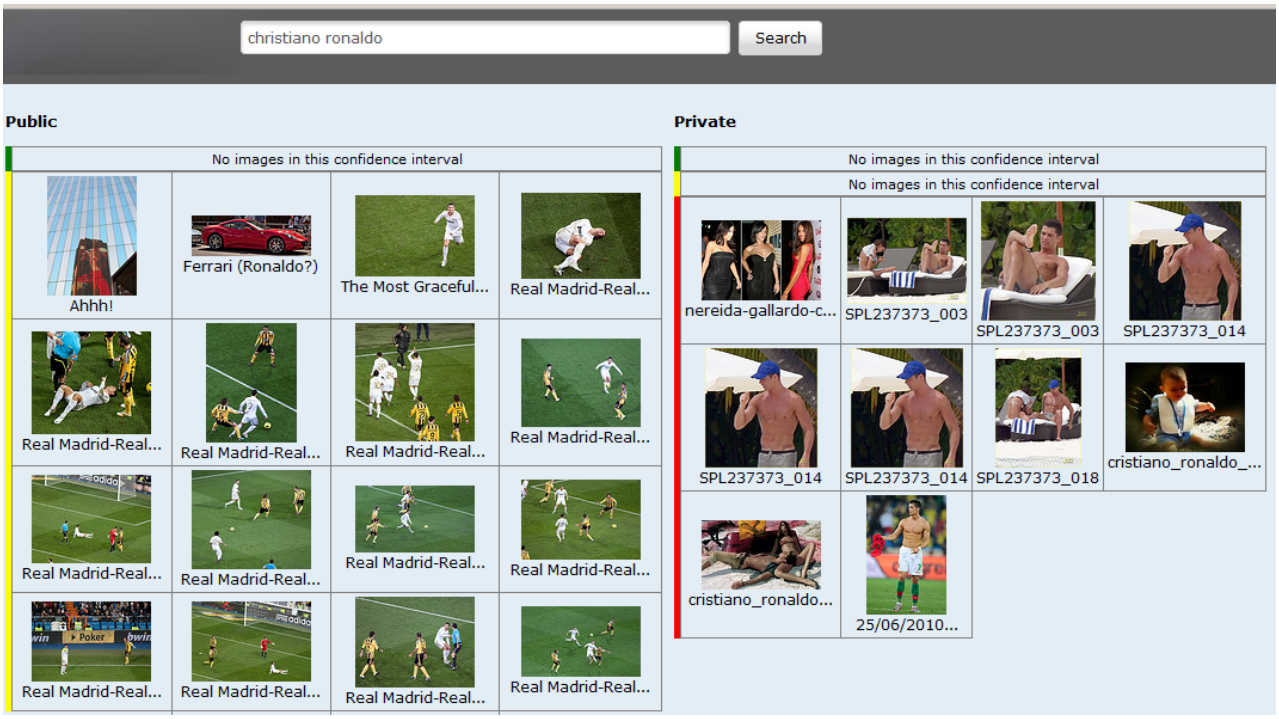

Figure 3: Private and public search results for the query "cristiano ronaldo" (June 06 2012).

2.4 Search

In order to create a list of images ranked by privacy, we estimated the likelihood of image privacy using the output of the SVM classifier trained on a set of images labeled as "public" or "private" by the users. We use the Flickr API as the underlying search provider for our PicAlert! search service. The user interface of the application simply consists of a text box and a keyword search can be performed pressing the "Search" button. The difference to other engines is mainly in the search result representation. PicAlert! divides the results into two sets: "public" and "private". Additionally, each set is divided into three subsets according to the classifier confidence intervals and is denoted by color. The green color corresponds to a strong classifier confidence, yellow to moderate and red to a weak confidence. Figure 3 shows an example for the results representation for the query "christiano ronaldo". In the left ("public") part we observe that the majority of pictures are related to sporting events, whilst the right ("private") part is mostly dominated by photos about Ronaldo's private life.

2.5 Classification GUI

The HTML based graphical user interface of the PicAlert! web service enables the user to submit images to the web service, and to obtain an estimate of the degree of privacy along with a visual explanation. The GUI consists of two main components - the image input page and the user setting page. The setting page allows users to obtain and to manage authorization keys for the web service client. The input page provides the following image input options:

- Direct Image URL: The service downloads the image available under the given URL. This image can be additionally supplied with title and tags.
- Flickr Image URL: The service downloads the image from the given Flickr URL. The service will also extract image title and tags if available.
- Image Upload: The user can upload an image directly from her desktop. Adding title and tags are optional.

The result page shown in Figure 2 appears after submitting the required data. The page contains the estimation of the image privacy value as the textual recommendation for the user to share the image or not. Additionally the features used for classification can be analyzed in more detail. Tables containing the most influential visual features are provided at the bottom of the page. If the user moves the mouse pointer over a particular feature name in the table, the corresponding feature is visualized within the image. The most discriminative image tags are also selected and presented to the user as a possible explanation for the classifier decision. Whilst the GUI currently only allows for the processing a single image at a time, batch image handling is possible through our XML web service interface.

3. SYSTEM EVALUATION

3.1 Classification Quality

In order to evaluate our classification approach, from the initial dataset we randomly sampled 60% as training data for building our classifiers, and 40% as test data, with each data set containing an equal proportion of public and private instances. Our quality measures for the classification are the precision-recall curves as well as the precision-recall break-even points for these curves. The break-even point (BEP) is equal to the F_1 measure and the harmonic mean of precision and recall. The results of the classification experiments for

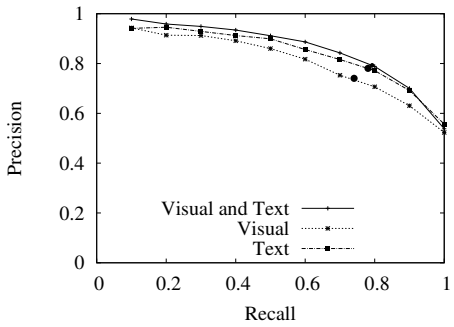

Figure 4: P/R curves for the features and their combination.

the visual features described in Section 2.2 and the combination of visual and textual features are shown in Figure 4. The *visual features* only lead to the BEP of 0.74. The *text features* provide a short but concise summary of the image content and result in a BEP of 0.78. Finally, the *combination of the visual and textual features* leads to an additional performance boost with a BEP of 0.80, showing that textual and visual features can complement each other in the privacy classification task. However, classification with only visual features alone also produces promising results, and is useful if limited or no textual annotations are available, as is the case for many photos on the web.

3.2 Search Quality

In order to evaluate our search ranking quality we randomly chose 50 image-related queries from an MSN search engine query log[1]. For each query, we computed privacy-oriented rankings using the pre-trained classifier. The list of test photos in descending order of their user-assigned privacy value was considered as ground truth for our experiments. We compared the order of the automatically generated rankings using Kendall's Tau-b [3]. We chose the Tau-b version in order to avoid a systematic advantage of our methods due to many ties produced by the high number of photos with equal user ratings. The original Flickr ranking does not consider the privacy of the images in the search results. This was reflected by a small τ_b value of -0.04. In contrast, our methods show a clear correlation with the user-based privacy ranking. The combination of textual and visual features provides the best ranking performance (τ_b=0.33).

4. DEMONSTRATION OVERVIEW

In the demonstration we will primarily show how the PicAlert! search system works and how the web service classifies images from different sources like personal desktops, mobile devices, Flickr, or arbitrary images from external URLs. We will demonstrate the graphical interface usage for private image search and upload. Additionally, we can elaborate in more detail on classification process and the visual feature analysis, and explain the XML web service interface.

Privacy Oriented Image Search Demonstration: Firstly the user types an arbitrary query into the search text field of the search interface. The results are separated into public and private sets according to the privacy value estimated by the classifier (Figure 3). We can observe that the proportion of the private results is strongly query dependent (e.g. "birthday" results in more private pictures than "tree"). The user can move the mouse pointer over a result image and see additional information including the privacy value computed by the classifier. Clicking on the image opens the corresponding Flickr page.

Classification Web Service Demonstration: We can select photos from our personal local file system (or memory sticks of visitors) and upload it to PicAlert! through the graphical upload control. The classification result is shown as a percentage value (Figure 2) with a text message. In order to visually support the user, traffic lights are displayed with colors corresponding to the classifier decisions. The user can move the mouse over the table with the visual features determined in her picture and the features are visualized by squares and stars determining the position and the rotation of the selected feature. The user can also repeat the process with a picture from Flickr, or an image denoted by an arbitrary URL.

Resources: Both demos are available on our web page[2] and can be used with any web browser. Additionally the user has a possibility to register and to obtain a key for the web service usage. A summary of the demo applications as well as a short video tutorial are available at the home page along with the annotated data described in Section 2.1.

5. CONCLUSIONS AND FUTURE WORK

The demo described in this paper introduces PicAlert! - a search engine for privacy oriented image search as well as web service for supporting user decisions regarding image privacy. This web service can be integrated into arbitrary social network and image sharing platforms, as well as browser plugins or mobile devices and prevent the user publishing potentially sensitive visual content about herself.

The popularity of mobile devices equipped with high definition cameras continues to increase. We plan to study pictures taken with mobile phones where we expect a larger proportion of private images. Apart from employing additional visual features, we further plan to include context recognition, which has become possible due to recently released powerful operation systems for smartphones and their sensors for GPS location, temperature, or acceleration.

6. REFERENCES

[1] S. B. Barnes. A privacy paradox: Social networking in the united states. *First Monday*, 11(9), Sept. 2006.

[2] T. Joachims. Making large-scale support vector machine learning practical. *Advances in kernel methods: support vector learning*, pages 169–184, 1999.

[3] W. H. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861, 1958.

[4] D. Lowe. Distinctive image features from scale - invariant keypoints. *IJCV*, 60(2):91–110, Jan. 2004.

[5] V. Schleswig-Holstein. Statistische erfassung zum internetverhalten jugendlicher und heranwachsender. In *A study of the consumer organization in Schleswig-Holstein, Germany*, March 2010.

[6] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova. Privacy-aware image classification and search. In *SIGIR 2012, Portland, USA*.

[1] http://research.microsoft.com/en-us/um/people/nickcr/wscd09/

[2] http://l3s.de/picalert/