Classifying Covert Photographs

Haitao Lang[1,2] Haibin Ling[2]
[1] Dept. of Physics & Electronics, Beijing University of Chemical Technology, Beijing, China, 100029
[2] Dept. of Computer & Information Sciences, Temple University, Philadelphia, PA, USA, 19122
langht@mail.buct.edu.cn, hbling@temple.edu

Abstract

The advances in image acquisition techniques make recording images never easier and brings a great convenience to our daily life. It raises at the same time the issue of privacy protection in the photographs. One particular problem addressed in this paper is about covert photographs, which are taken secretly and often violate the subjects' willingness. We study the task of automatic covert photograph classification, which can be used to help inhibiting distribution of such images (e.g., Internet image filtering). By carefully collecting and investigating a large covert vs. non-covert photographs dataset, we observed that there are many features (e.g., degree of blur) that seem to be correlated with covert photographs, but counter examples always exist. In addition, we observed that image visual attributes (e.g., photo composition) play an important role in distinguishing covert photographs. These observations motivate us to fuse both low level images statistics and middle level attribute features for classifying covert images. In particular, we propose a solution using multiple kernel learning to combine 10 different image features and 31 image attributes. We evaluated thoroughly the proposed approach together with many different solutions including some state-of-the-art image classifiers. The effectiveness of the proposed solution is clearly demonstrated in the results. Furthermore, as the first study to this problem, we expect our study to motivate further research investigations.

1. Introduction

1.1. Background

Proliferation of image/video acquisition devices and new internet technologies provide people great convenience to shoot photos and share them through websites such as Google Picasa[1] and Flickr[2]. Such convenience is however accompanied with the issue of privacy protection, which recently started drawing research attention in computer vision community [28, 4, 13, 8, 22]. In this paper, we investigate a new type of visual privacy threaten, named covert photography, in which the subject being photographed is unaware that he or she is intentionally photographed. Photos taken this way, named *covert photographs*, often seriously threaten public or personal privacy. For example, some people spy on neighbor's home activities with a night vision camera; paparazzi stalk celebrities to shoot pictures of their private life; voyeurs capture photos using hidden cameras in public restrooms, etc. Such photographs or videos often seriously jeopardize public privacy, and, when distributed on the Internet, can cause even worse consequences [17].

Many states and countries have enacted laws, regulations and policies to forbid or restrict inappropriate photographing activities and distribution of related photographs [32, 20, 15, 11]. For example, laws have been passed to prohibit photographing in privacy sensitive locations such as dressing rooms and restrooms [11]; the French "*Presumption of Innocence and Rights of Victims*" [32] legislation prohibits "*any publication of a person without their express consent*"; the United Kingdom enacted "*Human Rights Act 1998*" [20] to restrict on the publication of photography to protect the public privacy.

Content-based filtering, either automatic or semi-automatic, is an important tool for controlling the distribution of inappropriate images or videos over Internet. For example, Google SafeSearch is designed to filter out the adult-oriented text, video, and images. Image filtering technologies have recently been attracting increasing research attention in computer vision and data engineering communities. Most of such methods focus on classifying adult images. For example, the pioneer work in [10, 12] combines color and texture analysis for skin detection and then uses geometric constraints to group skin regions for naked body detection. In [19] a statistical skin color model is developed for adult images identification. In [18], contour-based visual features and text information are fused for recognizing pornographic web pages. Bag-of-features is used in [7] to classify pornographic images. More studies can be found in

[1] http://www.picasaweb.google.com/

[2] http://www.flickr.com/

[16, 27, 31], *etc*.

1.2. Our Contribution

Despite the large amount of previous studies on content based image filtering (e.g. those mentioned above), little effort has been devoted to handle covert photos. By our observation, there are at least two issues need to be addressed: First, covert photos recognition task is a bit more ethically challenging than pornography recognition. Pornography is more or less defined by the subject of the photograph, whereas covert photography is defined by the acquisition method. It is not easy to collect an annotated ground truth dataset. Second, classification of covert photos is not an easy task due to the large intra-class variations and the fact that there are many regular photos that resemble covert ones. Some example photos from the dataset we collected are shown in Figures 1 and 7.

In this paper, we study the problem of *covert photograph classification* by addressing both issues. We follow Wikipedia [37] to formally define the covert photographs. Guided by the definition, we collected an initial image dataset from multiple sources. By checking images in the collected dataset, we observed that both low level image statistics and middle level image attributes are relevant for distinguishing covert images from regular ones. Nonetheless, there are always exceptions for each individual image feature or attribute. Motivated by these observations, we propose to classify covert photographs by fusing information from both low-level image features and visual attributes. Specifically, we use multiple kernel learning (MKL) to combine 10 low level feature-based classifiers and 31 attribute-based classifiers. Finally, a thorough experimental evaluation is conducted using the collected dataset and our method demonstrates very promising performance.

Our major contribution lies in the study of automatic covert photo classification, which, to the best of our knowledge, has never been investigated before. Our contribution includes three interconnected components: (1) We have collected and annotated a covert–regular photos database containing 1500 covert photos and 6000 regular photos from varying sources, e.g., web, surveillance system, voyeurism publishing, real covert photography on street, and we plan to share it for research purposes. (2) We propose a new approach to combine image statistics and attributes for covert photo classification, which to the best of our knowledge has not been explored previously. (3) We evaluate the state-of-the-art image classification approaches on our database, including fine designed image descriptors, bag-of-features method, image statistics, and discriminative semantic attributes based methods. It is shown that our method clearly outperforms existing image classification methods for the covert photograph classification task.

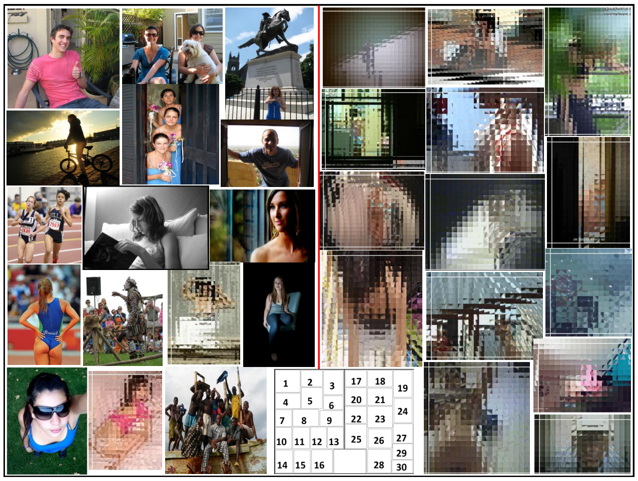

Figure 1. Sample photos from our covert database. Photos 1 to 16 on the left column are regular photos, while the remaining (17 to 30) are covert photos. All these photos can be classified correctly by our method, in the case of great inter-class similarity and intra-class variation. Note (for Figure 7 too): some images are made mosaic for the purpose of protection to kids and presonal privacy.

2. Problem Formulation

2.1. Problem Definition and Challenges

The problem of covert photograph classification is straightforward: given an input photo or image I, determine whether it is a *covert photo* or a *regular* one. So, when is a photo called "covert"? We use the definition of "secret photography", which is synonym of "covert photography", from Wikipedia as the answer [37]:

> "Secret photography refers to the use of an image or video recording device to photograph or film a person who is unaware that they are being intentionally photographed or filmed."

There are several similar or related concepts in photography regimes, such as *candid photography, voyeuristic photography, paparazzi*, *etc*. We use *covert photography* to emphasize the fact that photographed subjects are "unaware that they are being intentionally photographed", which more than often implies invasion of privacy.

From the definition, we can see two main challenges in the study of covert photograph classification. The first challenge lies in collecting a rigorous ground truth dataset for the investigation. On the one hand, it is almost impossible to trace the real acquisition method of each photo, which makes collecting a large database for research more difficult. On the other hand, it is inappropriate to collect such data by intentionally taking photo by ourselves, which may bring a serious bias. We handled this challenge by first collecting candidate photos from multiple sources and then decided whether they are covert by voting from human observers. The process is detailed in the next subsection.

The second challenge lies in the large inter-class similarity and inner-class variation. This challenge can be clearly seen in Figures 1 and 7, which show example covert photos in contrast to regular ones. This challenge makes it hard to use existing image classification algorithms that use a single type image features. For example, while many covert photos have low image qualities, there are some with qualities as good as regular ones. More detailed discussion is given in Section 3, where we describe our solution that combines low-level image features and middle-level image attributes.

2.2. Database Construction

There are billions of images on the Internet through image sharing website, however, due to the key words searching restriction, covert photos are not trivial to obtain. To collect a covert photo dataset for research usage, we seek image sources mainly through the following ways: 1) searching on the web with dozens of synonyms; 2) selecting frames from privacy invasion related surveillance videos; 3) querying from certain specific types of websites, such as peep tom, voyeurism blog, celebrity gossip; 4) submitted by volunteers for research purpose, e.g. photos captured on street *etc*. As we discussed in previous sections, whether an image is covert or not is defined by its acquisition method. When observing the raw covert dataset, we found most of them are truly taken secretly as indicated by the information provided by website or volunteers; but some of them do not have explicit notes regarding the photographing procedure. For this reason, we used human annotation for the decision. Ten human subjects (six males and four females) with ages ranging from 20 to 40 years volunteered to help. After explaining the definition of covert photography to them, we presented the initially collected images to them. A subject then labeled each image as either a covert photo or not. An image is treated as valid only if seven or more subjects agree on its labels. After the "raw data purifying" processing, 1,500 covert photos are finally determined from the 2,630 initially collected ones.

In addition to covert photos, regular photos as negative samples are collected from Flickr, Picasa Albums, and about 100 pictures from Caltech 256 "people" category. When collecting the dataset, an attention is paid to make it as diverse as possible, by including photos from various races, nationalities, ages, occupations, scenes, capture time, and professional or amateur photography *etc*. This collection contains 11,500 images initially. Then, after carefully checking the collection, we removed many (near) duplications and inappropriate ones (e.g. cartons). Finally, we kept 6,000 regular images in our dataset.

We noticed that some images in our collection have large caption areas, usually in the bottom region. To avoid bias, we manually cropped such images to remove such regions. No other preprocessing had been conducted on the images. In our experiments we split the datebase into training and testing sets. The training set contains 1,200 covert images and 4,800 regular images, while the testing set contains the remaining. Figures 1 and 7 show some samples from the database.

Table 1. Attributes and correlated low-level features. Note: Considering the limited space in tables and figures. We use abbreviations e.g. c-g, c-m, *etc*. instead of color gist, color moments, *etc*. The detailed correspondences between abbreviations and full names are listed in Section 3.

Group	**Attributes**	**Related features**
Image Quality	blur and noise	BIQI score [24]
Visualization	color richness image brightness color saturation contrast	c-g, c-m, hue c-g, c-m, g-hist hue, c-m glcm, g-hist
Image Contents	face presence human presence human dressing pornographic scenes time background	face detector [36] PHOG, e-hist, LBP c-g, c-m, hue c-m, hue c-g, LBP g-hist, e-hist e-hist, PHOG spatiogram
Photography	capture distance view angle	Dof [6] c-g, PHOG spatiogram

2.3. Attribute Annotation

Attribute information has recently been used in visual recognition problems [21, 29]. In order to use attributes for covert photograph classification, attribute labels have been collected which will be used in our task. The attributes in the dataset are selected from four groups, *i.e*. image quality, visual property, image content, and photography, as listed in Table 1. Ten volunteers are asked to annotate the attributes. To avoid the interference of prior knowledge about our final classification task to attribute annotation, the ten volunteers were selected to avoid any overlap with the volunteers for covert/regular labeling. Similar to the covert/regular labeling, after all subjects annotated attributes of all images, only those labels with super majority agreement (seven or more subjects) were kept.

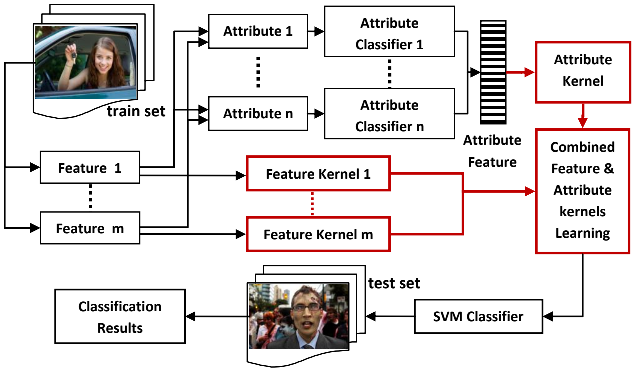

Figure 2. Framework overview.

3. Combine Image Features and Attributes

3.1. Overview

By investigating carefully the collected covert photo dataset, we observe that it is difficult to use single feature or attribute to distinguish covert photos from regular ones. This is due to the large inter-class similarity and inner-class variation, which are essential to the photography process of covert photos. In fact, as shown in subsection 4.3, classifiers based on individual image features have clearly rooms to improve.

Motivated by this observation, we propose to fuse information from multiple sources for classifying covert photos. In particular, our solution combines 10 different image features and 31 image attributes in a multiple kernel learning framework. The overview of the proposed method is illustrated in Figure 2. In the following subsections, we give details of each component in the framework.

3.2. Low-Level Image Feature

Many existing image classifiers use low level features directly extracted from images, such as GIST [26], local binary pattern (LBP) [1], histogram of oriented gradient (HOG) [5], or distribution of certain low level features based on various criteria, such as bag of features [9] *etc*., and have been proven to be promising in object recognition, image retrieval, category classification, *etc*. These features describe different image characteristics, such as holistic property [26], local image property [23], image patch characters [3], shape [26], color [33], spatial information [2], and some image statistical information [30]. Due to they all capture some discriminative information toward distinguishing covert photos from regular ones, it motivates us to fuse ten typical low level image features for our task, which described as follows:

- **Bag of Features** (BoF): We use SIFT [23] as the local descriptors and a vocabulary of size 180 to represent images.[3]

- **Color GIST** (c-g) [26]: We use a concatenation of the gist descriptor in the HSV color space. In every color channel, orientation histograms are computed on a 4×4 grid over the entire image. They are extracted at three scales with eight orientation bins on each scale. The final descriptor is a 1152 dimensional vector.

- **Color moments** (c-m) [30]: We concatenate the first three image moments, *i.e*. mean, variance, and skewness in Lab color space, over a 5×5 grid of entire image to construct a 225 dimensional descriptor.

- **Edge Orientation Histogram** (e-hist): We first extract Canny edges of the image, then compute five histograms, including four directional edge histograms (horizontal, vertical, two diagonals) and one non-directional edge, on three spatial pyramid levels (3×3, 4×4, and 5×5).

- **Gray Histogram** (g-hist): We simply compute the 256 gray level histogram of an image.

- **Grey Level Co-occurrence Matrix** (glcm): We scale the image to 16 gray levels, then calculate the occurrence frequency that a pixel with value i occurred in four-connection neighborhood of a pixel with value j, and return a gray-level co-occurrence matrix. Then we expand the matrix to a vector and create a 256 dimensional descriptor.

- **Hue descriptor** (hue) [33]: Since hue is known to be unstable around the grey axis, to make it more robust, we compute the hue histogram by weighting each hue value by its saturation to obtain a 36 dimensional descriptor.

- **Local Binary Pattern** (LBP) [25]: We compute the histogram of LBP code on a 4×4 grid over the entire image. The value of LBP code of a pixel is computed based on the binary number of its 8 surrounding neighborhoods within a circle with the radius of 1.

- **Pyramid histogram of orientation gradient** (PHOG) [3]: We first extract Canny edges, then quantize the gradient orientation on the edges (ranging from $0°$ to $360°$) into 40 bins. Three spatial pyramid levels, 1×1, 2×2 and 4×4, are used. The dimensionality of the final descriptor is 680.

- **Spatiogram** (spg) [2]: We quantize gray image into eight buckets, and then extract spatiogram histogram on three spatial pyramid levels (3×3, 4×4, and 5×5).

[3] For our task, we evaluate the vocabulary sizes from 100 to 300 and 180 performs best.

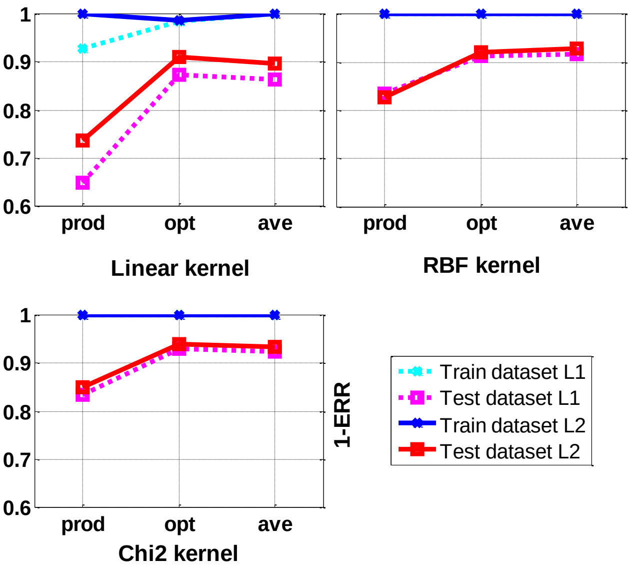

Figure 3. Accuracy evaluation of MKL methods

3.3. Attribute Classifiers and Attribute Feature

We list our attribute vocabulary in Table 1. For some attributes e.g. "blur and noise", "face presence", and "capture distance", we use the BIQI detector [24], the face detector [36] and the Dof detector [6] to calculate the scores directly for each image. For remaining, we use low level image features mentioned in previous subsection to train the attribute classifier by a supervised learning method for each attribute. For a given attribute, we use all ten low level features to train classifiers respectively and select those with the highest cross validation accuracies as the final attribute classifiers. For example, for attribute "color richness", there are three classifiers trained by color gist, color moment and hue descriptor respectively obtained the best performance. So, we kept these three attribute classifiers. By this way, we obtained in total 31 attribute classifiers, the final selected attributes and correlated training features are listed in Table 1. Then we construct an intermediate *"attribute feature"* by concatenating the prediction outputs of the classifiers to yield a 31-dimensional descriptor as depicted in Figure 2. In this phase, the role of "attribute feature" is same as low level features in the sense of representing an image. After "attribute feature" are obtained, we use standard feature kernel computation method to compute the attribute kernel.

3.4. Fusion with Multiple Kernels Learning

None of single feature have both highly invariant to the intra-class variations and powerful inter-class discriminative power to all classes. Recently, several methods have been proposed to combine multiple features to improve classification performance instead of using a single one.

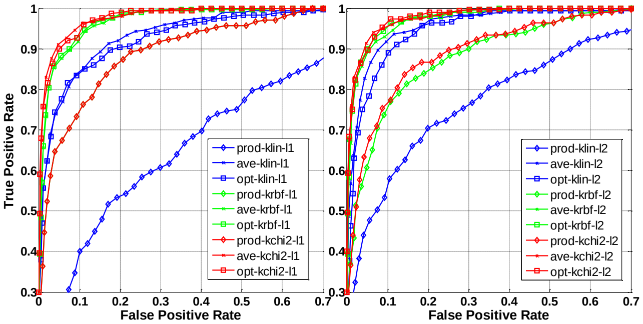

Figure 4. Performances of different configurations in MKL. For clearly illustration, we only show part of the curves which x and y in the range 0 to 0.7, and 0.3 to 1 respectively.

Table 2. Multiple kernels combination components

Normalization	ℓ_1-norm, ℓ_2-norm
Kernel Function	$k_{lin}(x_i, x_j) = \langle x_i, x_j \rangle$ $k_{RBF}(x_i, x_j) = \exp(-\gamma \|(x_i - x_j\|^2)$ $k_{\chi^2}(x_i, x_j) = \exp(-\gamma \frac{\|x_i - x_j\|^2}{x_i + x_j})$
Kernel Combination	ave: $\frac{1}{P}\sum_{m=1}^{P} k_m(x_i, x_j)$ prod: $(\prod_{m=1}^{P} k_m(x_i, x_j))^{1/P}$ opt: Equation (2)

Multiple Kernel Learning (MKL) [34, 35, 14] is one of successful methods. The core idea of MKL algorithm under the SVM framework is to seek optimal combination coefficients to kernel matrix, as shown in Equation (1):

$$\mathbf{K}_{opt} = \sum_{p=1}^{P} \eta_p \mathbf{K}_p \tag{1}$$

where P is the total number of kernels, η_p is the combination coefficient, and $\mathbf{K}_p$ is the kernel matrix. The element $k_p(x_i, x_j)$ of $\mathbf{K}_p$ defines the similarity between a pair of samples x_i and x_j. In our work, we solve an optimization problem as Equation (2)[34] to obtain combination coefficients:

$$\begin{aligned} &\min_{\mathbf{w},\eta,\xi,b} \ \tfrac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{Z} \xi_i + \sum_{p=1}^{P} \sigma_p \eta_p \\ &\text{s.t.} \ \ y_i(\langle \mathbf{w}, \Phi(x_i)\rangle + b) \geq 1 - \xi_i, \ \xi > 0, \ \eta > 0, \end{aligned} \tag{2}$$

where $\mathbf{w}$ is the vector of weight coefficients, C is the penalty parameter, ξ is the slack variables, b is the bias term of the separating hyperplane, Z is the total number of training image features, and $\Phi(x_i)$ corresponds to the feature space that implicitly constructs the combined kernel function $k(x_i, x_j) = \sum_{p=1}^{P} \eta_p k_p(x_i, x_j)$. To obtain the optimal combination coefficients effectively, we use a two-step

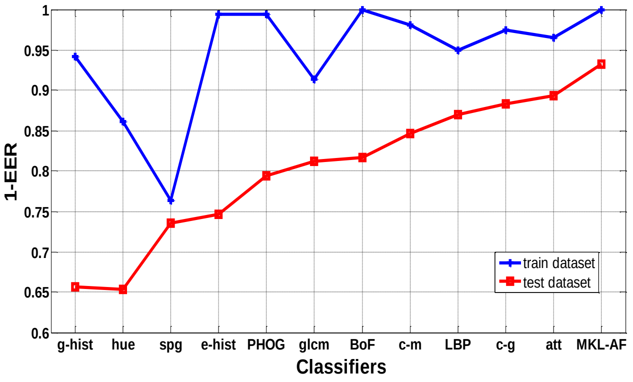

Figure 5. Performance evaluated by 1-EER. We ranked the classifiers based on the performance in test dataset. Label "att" and "mkl-a-f" represent visual attributes and combination of attributes and feature kernels by MKL (our method) respectively, for Figure 6 and Table 3 too.

SVM training method. At each iteration, we first update the combination coefficient η while fixing C, and we then update C while fixing η. These two steps are repeated until convergence.

Feature normalization and kernel standardization. Some attribute descriptors, come from attribute classifiers that can generate negative values, *e.g.* color richness attribute trained by color moments feature. This causes problems for some kernel functions, such as χ^2 kernel. Therefore, we feed feature and attribute descriptors into a sigmoid function, defined as

$$score_p = 1/(1 + e^{-score_{ori}}) \; , \tag{3}$$

to ensure the elements of the descriptors are non-negative. Normalization of descriptors is a trivial problem, but may have direct effect to classification performance [35]. In our work, we compare the performance of ℓ_1-norm and ℓ_2-norm based descriptors normalization. To standardize the entire kernel, we rescale it such that the variance $s^2 = \frac{1}{Z}\sum_{i=1}^{Z}(\Phi(x_i) - \overline{\Phi(x_i)})^2$ in the feature space remains const, which yield $K^* = K/(\frac{1}{Z}\sum_i K_{ii} - \frac{1}{Z^2}\sum_{i,j} K_{i,j})$.

4. Experiments and Discussion

4.1. Performance Evaluation Metrics

Considering the unbalance of our database, *i.e.* the ratio of covert and non-covert photos is 1 to 4 in both training and test database, we use AUC (area under curve) and 1-EER (1 minus equal error rate) of ROC curve to evaluate the performance, instead of classification accuracy.

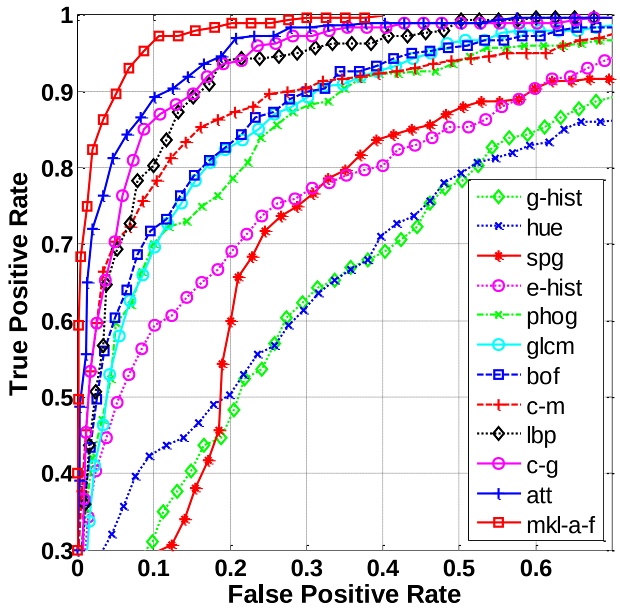

Figure 6. ROC Curves for evaluated methods. For clearly illustration, we only show part of the curves which x and y in the range 0 to 0.7, and 0.3 to 1 respectively.

4.2. Evaluation of MKL Algorithms

There are three important aspects to be considered for the combination of multiple kernels, *i.e.*, 1) the type of basic kernel functions; 2) the descriptor normalization method; and 3) the kernel combination method. In our experiments, for each component, we compare some typical choices. As listed in Table 2, we evaluate linear, RBF and χ^2 kernel functions, with either ℓ_1-norm or ℓ_2-norm normalizations. We also compare our optimization based kernel combining method (Equation (2)) with the two simple but empirically successful methods, *i.e.* summation averaging (ave) and multiplication averaging (prod). The results are illustrated in Figures 3 and 4. The results show that the performance of using ℓ_2-normalization is better than that of using ℓ_1-normalization. In our task, summation averaging performs similar to the optimization based method. Generally speaking, the method based on optimizing a kernel combination with the χ^2 kernel function yields the best performance. Thus, we use this combination to train our covert photographs classifier.

4.3. A Comprehensive Evaluation of Classification Methods on Covert Photographs Task

We compare our method with existing image and category classification methods on the covert database. We train classifiers on the training dataset and evaluate the performance on the test dataset. There are three types of classifiers in our evaluation: 1) those using classical single dis-

Table 3. AUC and 1-EER values

Method	g-h	hue	spg	e-hist	phog	glcm	BoF	c-m	LBP	c-g	att	mkl-a-f
AUC	0.7098	0.7198	0.7613	0.8204	0.8811	0.8938	0.9020	0.9092	0.9376	0.9487	0.9639	**0.9839**
1-EER	0.6567	0.6533	0.7358	0.7467	0.7942	0.8125	0.8167	0.8467	0.8700	0.8833	0.8933	**0.9325**

criminative low level image feature as described in Section 3.2, 2) the one trained by visual attributes (we use attribute vocabulary as introduced in Section 3.3), and 3) the proposed method. We plot the performances ordered by the 1-EER criterion in Figure 5 (the detailed scores are listed in Table 3). When considering the classifiers with only image statistics features, color gist, LBP, and BoF (SIFT) perform better than others. This is consistent to some degree with their performances shown in previous applications on visual classification. The good performance of color moments, however, surprises us and motivates us to study the power of image statistics on covert photograph classification. Visual attribute-based classifier outperforms discriminative feature-based classifiers, which shows that visual attributes provide important discriminative information for covert photographs. The detailed ROC curve and AUC data are shown in Figure 6 and Table 3 respectively. It shows clearly that the proposed method, by combining both image statistics and attributes, significantly outperforms all other methods.

Besides above experiments, we complete another interesting experiment to show how the performance of attribute classifiers affects the final covert image classifier. In the experiment, instead of using the trained attribute classifier, we use the ground truth attribute values directly for cover photograph classification. Specifically, we use the 14 attribute as shown in Table 1 to predict covert photographs. The result are 98.55% (cross validation accuracy) for training and 96.73% for test, which is about 3% higher than the proposed approach.

4.4. Discussions

We investigate the photos which are misclassified by the top six classifiers including ours, whose prediction accuracies (1-EER) are higher than 80% as show in Figure 5, and have two discoveries:

1) The incorrectly classified images by our classifier overlap largely with those by classifiers based on attributes, color gist, and color moments, In contrast, the overlap with errors from the BoF classifier is much smaller. As known, gist descriptor tends to extract the holistic characteristics of an image, and color moments by themselves are an image statistics feature. As for visual attribute, by checking the vocabulary, we find almost all of them reflect some global attribute of images. In comparison, BoF is a bundle of local image features (SIFT in our study). That is to say, the BoF classifier pays more attention to local details of images instead of holistic properties. Our classifier integrates more global features, makes its performance more similar to other global feature based classifiers than to the BoF classifier.

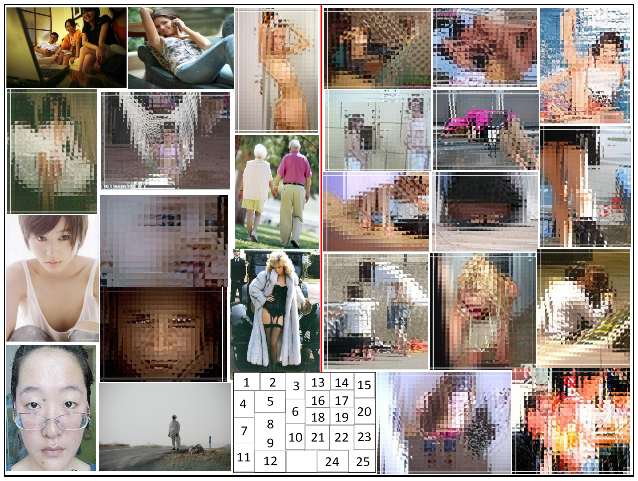

Figure 7. False classification samples by our method. Left(1-12): regular photos which are misclassified to covert ones. Right (13-25): covert photos which are misclassified to regular ones.

2) Most misclassified covert photos by our classifier (as shown in the right half of Figure 7) confuse other classifiers as well. By re-examining of the original sources of these images, we find that most of such images are labeled as “covert” by human volunteers, since the original sources do not provide related textual description. In other words, these images show the disagreement between human and the algorithms tested in the paper. This also suggests the importantance of using the “true ground truth”, instead of human annotation, in the future study.

5. Conclusions

In this paper, we introduce and study a novel image recognition/classification task, *i.e.* covert photograph classification. Comparing with the existing pornography/naked photograph recognition task, covert photograph classification is more challenging. Pornography is defined by subject of the photograph, whereas covert photography is defined by the acquisition method. It is more difficult to learn clues which might reflect images acquisition method than to learn those based on image contents only. We construct a large covert database for research purposes. By carefully investigating the similarities and differences between covert and regular photographs, we propose to fuse both low level images statistics and middle level attribute features using the MKL algorithm for classifying covert photos. We evaluate our method together with many modern image classifiers. The experimental results demonstrate our method outperforms all other competitors. In future study, we will construct a rigorous ground truth database. We are also interested in investigating more complementary low level image features and middle level image attributes, as well as other multi-feature combination approaches.

Acknowledgements. We thank Liang Du, Yi Wu, Erkang Cheng for many helpful discussions; Dr. Larry Davis and Dr. Jie Yang for stimulating comments; and reviewers for valuable suggestions. We are also grateful to volunteers who helped with data collection and annotation. The work is supported in part by NSF China (NSFC No.60803063) and in part by NSF (IIS-1049032).

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *PAMI*, 28(12):2037–2041, 2006. 4

[2] S. Birchfield and S. Rangarajan. Spatiograms versus histograms for region-based tracking. *CVPR*, 2005. 4

[3] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *CIVR*, 2007. 4

[4] D. Chen, Y. Chang, R. Yan, and J. Yang. Tools for protecting the privacy of specific individuals in video. *EURASIP J. Appl. Signal Process.*, 75427, 2007. 1

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *CVPR*, 2005. 4

[6] R. Datta, D. Joshi, J. Li, and J. Wang. Studying aesthetics in photographic images using a computational approach. *ECCV*, 2006. 3, 5

[7] T. Deselaers, L. Pimenidis, and H. Ney. Bag-of-visual-words models for adult image classification and filtering. In *ICPR* , 2008. 1

[8] L. Du and H. Ling. Preservative license plate de-identification for privacy protection. In *ICDAR*, 2011. 1

[9] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005. 4

[10] M. Fleck, D. Forsyth, and C. Bregler. Finding naked people. *ECCV* , 1996. 1

[11] The National Center for Victims of Crime. Video Voyeurism Laws, 2009. 1

[12] D. Forsyth and M. Fleck. Automatic detection of human nudes. *IJCV*, 32(1):63–77, 1999. 1

[13] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, L. Vincent. Large-scale privacy protection in google street view. *ICCV*, 2009. 1

[14] M. Gönen and E. Alpaydin. Multiple kernel learning algorithms. *JMLR*, 12:2211–2268, 2011. 5

[15] Public Law No: 108-495. Video Voyeurism Prevention Act of 2004. 1

[16] M. Hammami, Y. Chahir, and L. Chen. Webguard: A web filtering engine combining textual, structural, and visual content-based analysis. *IEEE TKDE*, 272–284, 2006. 2

[17] A. Hargrave and S. Livingstone. *Harm and offence in media content: a review of the evidence*. Intellect Ltd, 2009. 1

[18] W. Hu, O. Wu, Z. Chen, Z. Fu, and S. Maybank. Recognition of pornographic web pages by classifying texts and images. *PAMI*, 1019–1034, 2007. 1

[19] M. Jones and J. Rehg. Statistical color models with application to skin detection. *IJCV* , 46(1):81–96, 2002. 1

[20] http://www.legislation.gov.uk. Human rights act 1998. 1

[21] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable Visual Attributes for Face Verification and Image Search. *PAMI*, 2011. 3

[22] F. Li, Z. Li, D. Saunders, and J. Yu. A theory of coprime blurred pairs. *ICCV,* 2011. 1

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

[24] A. Moorthy and A. Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, 2010. 3, 5

[25] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 971–987, 2002. 4

[26] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV* , 42(3):145–175, 2001. 4

[27] H. Rowley, Y. Jing, and S. Baluja. Large scale image-based adult-content filtering. In *Int'l Conf. on Computer Vision Theory and Applications*, 2006. 2

[28] A. Senior. Protecting Privacy in Video Surveillance. In *Privacy Protection in a Video Surveillance System,* 2009.1

[29] B. Siddiquie, R. Feris, and L.S. Davis. Image Ranking and Retrieval Based on Multi-Attribute Queries. *CVPR,* 2011. 3

[30] M. Stricker and M. Orengo. Similarity of color images. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 2420:381–392, 1995. 4

[31] H. Sun. Pornographic image screening by integrating recognition module and image black-list/white-list subsystem. *IET Image Processing*, 4(2):103–113, 2010. 2

[32] Daily Telegraph, 28.08.05. The notoriously strict privacy laws in France ensure that such intrusion into the private lives of public figures is rare, 2005. 1

[33] J. Van De Weijer and C. Schmid. Coloring local feature extraction. *ECCV*, 2006. 4

[34] M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In *ICCV*, 2007. 5

[35] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *ICCV*, 2009. 5, 6

[36] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004. 3, 5

[37] http://en.wikipedia.org/w/index.php?title=Secret_photography &oldid=465418243. Secret photography Wikipedia, the free encyclopedia, 2011. 2