ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 1 January 2016, Page No. 15592-15597

A Survey on Covert photo Classification.

Vrushali C Bhadane[1], N.K.Zalte[2]

1 M.E Student, Kalyani Charitable Trust's Late G.N. Sapkal College of Engineering, Nashik

2 Assistant Professor, Kalyani Charitable Trust's Late G.N.Sapkal College of Engineering, Nashik

Vrush.c.b@gmail.com, nzalte@gmail.com.

Savitribai Phule Pune University

Abstract

The advances in image acquisition technique made 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 subject willingness. Covert photos are often privacy invasive and, if distributed over Internet, can cause serious consequences. Automatic identification of such photos, therefore, serves as an important initial step toward further privacy protection operations. The problem is, however, very challenging due to the large semantic similarity between covert and noncovert photos, the enormous diversity in the photographing process and environment of cover photos, and the difficulty to collect an effective data set for the study. To overcome these challenges three contributions are used. First is to consider a dataset of 2500 covert images and verify each image carefully. Second is to conduct user study on how human perform to distinguish between covert and non-covert images. Third is to perform covert photo classification algorithm that fuses various image features and visual attributes in the multiple kernel learning framework.

Index Terms *— Privacy protection, covert photography, image classification, visual attribute, multiple kernel learning.*

I.Introduction

Proliferation of image acquisition devices and new Internet technologies provide people great convenience to shoot photos and share them publicly. Such convenience is however accompanied with a series of social and legal issues, Such as adult contents controlling and privacy protection. They make use of a new type of visual privacy threat to investigate, called covert photography, in which the photographing process is intentionally concealed from the subject being photographed. Photos taken this way, named covert photographs or covert photos, often seriously threaten public or personal privacy.

Despite a large literature on image understanding, covert photo classification has never been studied to the best of our knowledge and it challenges existing image classification methods from several aspects. First, there is a large ambiguity between covert and non-covert photos in terms of semantic content. This is because covert photos distinguish themselves by their acquisition procedure other than semantic content. Second, there is a large variation in the photographing environment and conditions for covert photos, and the variation causes serious within-class variation of covert photos. Third, there exists no covert photo dataset and it is nontrivial to collect one. For addressing these challenges three contributions are made towards automatic covert photo classification.

• First, a large covert photo dataset containing 2,500 covert photos and 10,000 non-covert ones is created. Each covert or non-covert photo is verified by checking its photographing process. The datasets are carefully designed so as to reduce potential biases as much as possible.

• To help for better understand the problem, they conduct a user study to investigate how human beings distinct covert photos from non-covert ones. The study generates an important benchmark measuring human's ability to the task.

Covert photo classification includes three aspects for classification as: Image representation, Attribute analysis and Feature fusion. For image representation a popular representation i.e. bag-of-features is used, where statistics collected from quantized local image descriptors such as SIFT are used to represent an image. Another powerful image/scene descriptor is the GIST descriptor that captures the global layout of the scene by describing the spatial distribution of textures. Visual attributes typically capture middle-level descriptive information. A notable application of visual attributes is face recognition, where the recognition performance is based on integrating attributes such as gender and age. In semantic attributes (e.g., "spotty") are used to describe familiar objects (e.g., "dog") for recognizing unseen objects with limited instances.

Combination of heterogeneous visual features and/or attributes had been investigated for visual recognition. In this method they follow the recently proposed strategy of multiple kernel learning (MKL)framework. Intuitively, MKL can be viewed a kernel fusion technique which treats different feature channels as different kernels and fuses heterogeneous information via kernel combination.

II. Related Work:

In 2001, IMAX query processing originates from influence maximization. Domingo's and Richardson [4] first study influence maximization as an algorithmic problem based on a Markovrandom field. One of the major applications of data mining is in helping companies determine which potential customers to market to. If the expected profit from a customer is greater than the cost of marketing to her, the marketing action for that customer is executed. So far, work in this area has considered only the intrinsic value of the customer (i.e, the expected profit from sales to her). We propose to model also the customer's network value: the expected profit from sales to other customers she may influence to buy, the customers those may influence, and so on recursively. Instead of viewing a market as a set of independent entities, we view it as a social network and model it as a Markov random eld. We show the advantages of this approach using a social network mined from a collaborative ltering database. Marketing that exploits the network value of customers |also known as viral marketing |can be extremely effective, but is still a black art. Our work can be viewed as a step towards providing a more solid foundation for it, taking advantage of the availability of large relevant databases.

In 2003 ,Since influence maximization is NP-hard, Kempe et al. propose a greedy method and show that its accuracy is higher than those of other naive methods. [1] represented an essential algorithmic issue for such informal community forms: in the event that we can attempt to persuade a subset of people to receive another item or development, and the objective is to trigger a huge course of further selections, which set of people would it be advisable for us to target? The enhancement issue of selecting the most influential hubs is NP-hard here, and we give the first provable estimate ensures for efficient calculations. Utilizing an investigation system in view of submodular capacities, we demonstrate that a characteristic avaricious technique acquires an answer that is provably inside 63% of ideal for a few classes of models; our system proposes a generalmethodology for thinking about the execution assurances of calculations for these sorts of influence issues in informal communities. We additionally give computational trials on substantial cooperation systems, demonstrating that notwithstanding their provable assurances, our estimation calculations significantly out-perform node selection heuristics taking into account the all around examined thoughts of degree centrality and separation centrality from the field of information.

In 2007,Leskovec et al.[5] improve the greedy method with the Cost-Effective Lazy Forward (CELF) selection [5].Given a water dispersion system, where would it be advisable for us to place sensors to rapidly distinguish contaminants? On the other hand, which online journals would it be advisable for us to peruse to abstain from missing essential stories? These apparently different issues offer basic structure: Outbreak discovery can be demonstrated as selecting hubs (sensor areas, web journals) in a system, keeping in mind the end goal to identify the spreading of an infection or data as fast as could be allowed. We exhibit a general strategy for close ideal sensor arrangement in these and related issues. We show that numerous sensible flare-up discovery targets (e.g.,detection probability, populace affected) display the property of "submodularity". We misuse submodularity to build up an efficient calculation that scales to substantial issues, accomplishing close ideal positions, while being 700 times speedier than a straightforward insatiable calculation. We additionally infer online limits on the arrangements' nature acquired by any calculation. Our calculations and limits likewise handle situations where hubs (sensor areas, web journals) have different costs. We assess our methodology on a few extensive certifiable issues, including a model of a water dissemination system from the EPA, and genuine web journal information. The acquired sensor situations are provably close ideal, giving a consistent portion of the ideal arrangement. We demonstrate that the methodology scales, accomplishing speedups and funds away of a few requests of greatness. We likewise demonstrate how the methodology prompts more profound bits of knowledge in both applications, noting multicriteria exchange off, cost-affectability and speculation.

In 2009,Chen et al. [8], [9] focus on reducing the cost for calculating the influence spread. They propose a greedy method based on randomly generated graphs and a degree-based method wherein the largest effective degree nodes are selected as influentialseeds. They also propose prefix excluding maximum influence arborescence (PMIA) heuristics where seed nodes influence the other nodes along the maximum influence path from a seed node to each node [9]. In the PMIA heuristics, if the maximum influence path from seed node s to node v includes another seed node s0 in their greedy-based algorithm, then their algorithm calculates the next maximum influence path from s to v which does not include s0. However, since calculating it in query processing time is expensive, the PMIA heuristics are inefficient for IMAX query processing. As the PMIA heuristics, the proposed method in this paper also uses such maximum influence paths, but it is more efficient than the PMIA heuristicsbased on keeping multiple alternative paths on a novel preprocessed structure.

In 2010,Wang et al. [7] in this paper we propose another calculation called Communitybased Greedy calculation for mining top-K influential hubs. The proposed calculation envelops two segments: 1) a calculation for taking so as to distinguish groups in an informal community into record data dissemination; and 2) a dynamic programming calculation for selecting groups to find influential hubs. We additionally give provable estimation sureties to our calculation.

Exact studies on an expansive genuine versatile interpersonal organization demonstrate that our calculation is more than a request of sizes speedier than the best in class Greedy calculation for finding top-K influential hubs and the blunder of our approximate algorithm is small.

In 2011 Jiang et al. [10] present simulated annealing-based methods that are used to escape the confinement problem of the greedy approach. The issue of influence boost, i.e., mining top-k influential hubs from an interpersonal organization such that the spread of influence in the system is amplified, is NP-hard. The greater part of the current calculations for the issue depend on insatiable calculation. Albeit eager calculation can accomplish a decent estimation, it is computational costly. In this paper, we propose a very surprising methodology in light of Simulated Annealing(SA) for the influence amplification issue. This is the first SA based calculation for the issue. Also, we propose two heuristic systems to quicken the merging procedure of SA, and another technique for registering influence to accelerate the proposed calculation. Trial results on four genuine systems demonstrate that the proposed calculations run quicker than the cutting edge covetous calculation by 2-3 requests of greatness while having the capacity to enhance the exactness.

In 2011,Kempe et al.(KKT) demonstrated the issue of influence amplification is NP-hard and a straightforward ravenous calculation ensures the best conceivable estimation component in PTIME. In any case, it has two noteworthy wellsprings of inefficiency. To begin with, finding the normal spread of a hub set is NP-hard. Second, the fundamental ravenous calculation is quadratic in the quantity of hubs. The first source is handled by evaluating the spread using so as to utilize Monte Carlo reenactment or heuristics. Leskovec et al.[6] proposed the CELF calculation for handling the second. In this work, we propose CELF++ and observationally demonstrate that it is 35-55% quicker than CELF.

III. Analysis of Our Work:

The proposed structure mainly focuses on following areas:

In this work, in order to give a new algorithm for the Influence Maximization problem, that keeps into consideration the history of actions that have been taken by the users in determining their influence over each other. Also, it uses the concept of community detection and its relationship with the field of Viral Marketing. I propose that instead of the Social Marketing &Influence model which has been used to simulate propagation of influence.

Module 1 :Social Activities Analysis

The algorithm starts by assigning credits to users based on their activities as shown in the log. Here is an example as to how this would work. Our Web Portal is one of the main players in the social activities. Here, an action is a user rating a edge for movement or activity. In other words, if user v rates “post”, and later on v is friend u does the same, it would be considered that the action of rating “post” has propagated from v to u. Although unlike the actual algorithm, which is not a propagation algorithm.

Module 2: Topic influence Analysis

Now, to use this scanning of action log to determine probabilistic influence between any two users. Once we have these influence values we will apply topic aware influence maximization framework along with linear threshold model so that performance and influence result should get improved, with probability values that are actually significant. This approach is clearly more practical and hence more accurate than assigning random probability values to each of these edges.

Module 3: Result Publish Phase.

In this stage, system will suggest influenced topic to their respective audience or users.

Mathematical Model:

Mathematical Model for Proposed Work

Assumptions:

S: System; A system is defined as a set such that:

S = {I, P, O}.

Where,

U: Set of users

= {UR: Set of Registered Users,

UN: Set of Un-Registered Users}

I: Set of Input.

O: Set of output.

P: Set of Processes.

INPUT SET DETAILS:

PHASE 1: REGISTRATION.

Ir= { username: i1,

Address: i2,

Pincode: i3,

Mobile no: i4,

Email: i5,

Images: i6,}

PHASE 2: Influence Maintenance & Maximization

Iv= { username: i1,

postinfo: i2,

commentdetalis: i3,

sharedetails: i4,

CandidateList: i5}

PROCESS SET DETAILS:

PHASE 1: REGISTRATION.

P1={ User registration: p11}

PHASE 2: Influence Maintenance & Maximization

P2={postdetails: p21,

Rating:p22,

commenting: p23

greedymethod: p24,

analysis : p25}

PHASE 3: Result

P3={SR_Statistic : p31,

SR_Result : p32}

OUTPUT SET DETAILS:

PHASE 1: REGISTRATION.

O1={userid: o11,

Password: o12}

PHASE 2: Influence Maintenance & Maximization

O2={PostClassification: O21,

infoRecommendation: O22,

influence_maximization: O23}

PHASE 3: Result

O3={DR_Statistic : o31,

DR_Result : o32}

IV. Conclusion:

In this paper, we detail IMAX question preparing to augment the influence on specific clients in informal organizations. Since IMAX question handling is NP-hard and computing its target capacity is #P-hard, we concentrate on the most proficient method to rough ideal seeds efficiently. To surmise the estimation of the goal capacity, we propose the IMIP model taking into account freedom between ways. To process an IMAX question efficiently, extricating possibility for ideal seeds is proposed and the quick avaricious based guess utilizing the IMIP model. We tentatively exhibit that our recognizing nearby influencing areas system is powerful and the proposed strategy is for the most part no less than a request of extent quicker than PMIA and IRIE with comparable exactness what's more, the proposed system is generally six requests of size speedier than CELF++ and the distinguishing neighborhood influencing districts method makes CELF++ around 3.2 times quicker while accomplishing high accuracy.

Later on, for IMAX inquiry handling, we will consider more different circulations of targets, for example, clients in the same group or the same college in view of the static profiles of clients. Next, we will apply IMAX question preparing to the straight limit model, and test whether the thoughts in this paper are still appropriate.

References

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proc. 9th ACM SIGKDD Int. Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038.

[8] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 199–208.

[10] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," inProc. 25th AAAI Conf. Artif. Intell., 2011, pp. 127–132.

Conf. Knowl. Discovery Data Mining, 2003, pp. 137–146.

[2] F.-H. Li, C.-T.Li, and M.-K. Shan, "Labeled influence maximization in social networks for target marketing," in Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust, Int. Conf. Social Comput., 2011, pp. 560–563.

[3] W. Lu and L. Lakshmanan, "Profit maximization over social networks," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 479–488.

[4] P. Domingos and M. Richardson, "Mining the network value of customers," in Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2001, pp. 57–66.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2007, pp. 420–429.

[6] A. Goyal, W. Lu, and L. V. Lakshmanan, "CELF++: Optimizing the greedy algorithm for
[7] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining

influence maximization in social networks," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 47–48.

[9] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in Conf. Knowl. Discovery Data Mining, 2010, pp. 1039–1048