CAREER: Observing the world through the lenses of social media

This is a five-year proposal for an integrated research, education, and outreach program on the algorithms and technologies needed for data mining in large, unstructured collections of photographs from online social media. The billions of photos on these sites, including both their visual content and non-visual metadata like text tags, GPS coordinates, and timestamps, contain an enormous amount of latent information about the state of the world and about human activities. The proposed project is based on the hypothesis that if aggregated together, these billions of observations could create virtual 'distributed cameras' or 'social sensor networks' that produce fundamentally new sources of observational data about the world, with the potential to impact a wide range of scientific disciplines that require such data including ecology, biology, psychology, and sociology. Unlocking this latent information will require innovative algorithms and technologies in data mining, applied machine learning, and large-scale computer vision.

The proposed research plan involves both investigating these core technologies and validating them with applications in ecology and psychology through interdisciplinary collaborations. The PI's expertise and recent work in modeling social networks, mining social media, recognizing objects and scenes, and reconstructing 3D models from large-scale photo collections lays the foundation for this project. The proposed education plans will develop new curricula for data mining and computer vision at both the undergraduate and graduate levels to help prepare students for the era of 'big data' jobs that increasingly rely on machine learning, data mining, probabilistic inference, and statistics. An outreach plan will increase undergraduate involvement in research (particularly for female and minority students) through summer research experiences, will educate general audiences about the challenges and opportunities of data mining and social media, and will help to connect the data mining and computer vision communities through tutorials and workshops.

Intellectual Merit: The research plan will investigate the feasibility of using large-scale social image collections for automated observation of the world. The plan will create new algorithms for visual social media mining by combining analysis of both visual evidence in photographs and non-visual metadata. Statistical and learning-based approaches will be investigated to understand and mitigate the effects of noise and bias in making accurate crowd-sourced observations. Innovative algorithms that leverage large-scale data to improve classic computer vision problems like scene recognition and 3d reconstruction will be investigated. These techniques will be validated on applications from biology, sociology, and ecology, comparing observational estimates produced by social media with actual ground truth data to produce quantitative assessments of accuracy and to characterize advantages and limitations of these approaches.

Broader Impact: The proposed research plan has the potential to create fundamentally new sources of observational data for a variety of scientific disciplines, which will be validated through interdisciplinary collaborations. Within Computer Science, the plan helps to bridge the fields of computer vision and data mining, which have traditionally been disconnected, through joint publications, a web resources portal, and by organizing tutorials and workshops about visual mining at both computer vision and data mining conferences. The education plan will prepare students for the next generation of technical jobs associated with a 'big data' world through a new graduate course in machine learning and a new undergraduate course in computer vision, and will distribute innovative materials for an undergraduate web search and data mining course on the web. Through partnerships with the Alliance for the Advancement of African-American Researchers in Computing, the National Center for Women in Technology, and Historically Black Colleges and Universities (HBCUs), the outreach plan will recruit students from minority and other under-represented groups to the PI's undergraduate and research programs. An annual workshop with the IU Lifelong Learning program will educate general audiences, particularly senior citizens, about data mining and social media.

1 Overview and Vision

Every day, millions of people across the world take photos and upload them to photo-sharing websites. Their goal is to share photos with friends and others, but collectively they are creating vast repositories of visual information about the world and its people. Each photo is an observation of *how the world looked at a particular point in space and in time*, as well as a record of *where the photographer was and what he or she was paying attention to*. Aggregated together, these enormous collections of observations — almost 150 billion photos just on Facebook and Flickr alone, or nearly 4% of the estimated 3.5 trillion photos taken in all of human history [154, 186] — could lead to revolutionary new ways of collecting observational data about the state of the world and human behavior, impacting a range of disciplines including ecology, biology, sociology, and psychology. This proposal is a 5-year Career plan that integrates research, teaching, and outreach activities in order to begin to make this vision a reality. The research objective of the project is to investigate how to analyze vast collections of user-contributed images and noisy metadata (including geo-tags, timestamps, text tags, and captions) in order to make accurate estimates about the state of the world, by combining research that spans both data mining and computer vision.

As a motivating example in support of this vision, consider the many biologists and ecologists that study the effects of a changing climate on the populations, distributions, and lifecycles of flora and fauna [145]. Monitoring these changes is surprisingly difficult: plot-based studies involving direct observation of small patches of land yield high-quality data but are costly and possible only at small scales [144], while aerial surveillance gives data over large land areas, but only certain types of ecological information can be observed from the air, and even these can be obscured by clouds, trees, and atmospheric conditions [39, 148]. But vast online social photo collections present an entirely new source of data: a large fraction of photos contain visual evidence about the environment, either on purpose (e.g., when people photograph birds and flowers) or incidentally (e.g., when animals or weather phenomena are visible in the background of a scene). By recognizing this visual evidence, combining it with metadata like photo geo-tags and timestamps, and aggregating observations from millions of people, revolutionary new ground-level, continental-scale datasets of the natural world and how it is changing over time could be created.

As a second motivating example, consider the scientists in a range of disciplines, from psychologists and sociologists to architects and urban planners, who study the preferences and behaviors of people, including how they navigate physical space [96] or how visual preferences vary across genders and cultures [97, 140]. Most existing work in these areas study small groups of people in laboratory settings. Social photo-sharing sites provide a vast new source of data for these studies, because they record where millions of people were located and what they were paying attention to as they go about their lives in the real world.

The proposed project is part of of the young but growing trend of mining 'big data' to produce fundamentally new paradigms for observation [13, 137, 147], combining together evidence produced by the activities of online users to create unstructured, opportunistic networks of 'social sensors' to observe the world [10]. Particularly striking examples include work showing that the spread of flu can be tracked via search queries [83], that earthquake activity can be monitored in real time via Twitter [67, 160], that online data can be used to answer fundamental questions in sociology [114], that social media can monitor the aggregate mood of humanity [87], and that analysis of historical newspaper and book articles can quantify changing cultural trends [132]. Nearly all of this existing work has used textual analysis (of documents, query terms, and microblogs), which has limited the types of sensing that can be accomplished. The proposed project is based on the hypothesis that analyzing visual data could lead to even more striking applications, because an image can record a much larger space of observational information than a tweet or query term, and visual information provides stronger evidence that can be verified in ways that simple text reports cannot.

With this promise comes significant challenges, however. Perhaps chief among these is that visual information is evidence in a 'rawer' form: text, unlike an image, is semantic meaning that has already been translated into human language. Extracting semantics from images is a notoriously difficult problem with a nearly 50 year history in the computer vision community [174]. Fortunately, recent work by the PI and others has shown that although vision is imperfect on individual images, some problems are (counter-intuitively) easier on large social image collections, for two main reasons: (1) these images generally includes a variety of non-visual metadata, including user-contributed comments, captions, and text tags, as well as camera metadata like timestamps, geo-tags, and EXIF camera parameters, that ease the semantic extraction problem by providing (noisy) evidence in addition to the visual information itself, and (2) social media collections contain a large degree of redundancy (e.g. many photos of the same place taken by different people), allowing robust estimation algorithms to produce good results even if a large minority of photos are discarded due to perception failures. Recent impressive work on social media collections, including large-scale 3D reconstruction [54, 55, 171], scene completion [93], and geolocation [44, 94, 119], suggests that our vision of large-scale visual social sensing is possible without solving the general computer vision problem.

Specific challenges. Four particular threads of investigation are needed to realize visual social sensing:

1. *Calibrating the social sensors,* and in particular knowing accurately where and when each photo was taken. Modern cameras record timestamps in the EXIF metadata, but these timestamps are often inaccurate (because users do not set camera clocks correctly or forget to reset them when they traverse time zones). Many social photos have geo-tags, but these are also quite noisy because they come either from hand-labeling or from imprecise consumer GPS receivers [20].
2. *Extracting semantic evidence from images,* using both visual analysis as well as non-visual cues that are available on social sharing websites, including text tags, captions, comments, and EXIF metadata (like focal length, focal distance, etc.). Analyzing visual content is challenging because computer vision is a very difficult problem, while non-visual data is easier to analyze algorithmically but is often very noisy (e.g. text tags that are ambiguous or have been added accidentally).
3. *Aggregating observations,* combining evidence from individual photos to produce accurate, coherent estimates of the state of the world. The challenges here are various sources of bias, including geographic bias caused by the uneven geographic distribution of photographic activity [44], and observer bias due to the fact that social media users are often younger and more tech-savvy than the general population. Moreover, observations are incomplete: the fact that no one took an image of an object at a particular point in time and space does not necessarily mean that it was not there, for example.
4. *Applying to important problems* in need of observational data, to quantify the accuracies and limitations of the proposed data mining techniques as a source of data about the world.

The proposed 5-year Career plan will address each of these four research threads (calibration, extraction, aggregation, and application) in order to work towards making large-scale visual social sensing a reality.

Contributions. The specific contributions of this 5-year Career plan are:

1. Creating techniques for precise geo-location and timestamping of observations;
2. Characterizing and alleviating bias in social media observations;
3. Developing new approaches for image understanding using metadata analysis, scene classification and object recognition;
4. Investigating the performance of these techniques to mine social media photo collections for world-scale observation, and testing them on several interdisciplinary applications;
5. Integrating research with education for both undergraduate and graduate students;

6. Working to attract women and minorities to computing;
7. Developing new curricula to prepare students for the skills needed in the Big Data era; and
8. Working to educate general audiences about social media and data mining.

PI qualifications. The PI brings the unique combination of expertise in data mining, social media analysis, and computer vision necessary to conduct the proposed research plan. The PI's work in computer vision has created algorithms for video indexing [15, 42, 51, 82], facial image detection and enhancement [124], human pose estimation [64], landmark recognition [119], and object recognition [46, 47, 49, 50, 52, 53, 65, 124, 125]. The PI's Ph.D. thesis [41] introduced new models and algorithms for recognizing broad object categories like cars, airplanes, and televisions; today's *de facto* standard recognition algorithm uses this same basic framework [71]. The PI's data mining work has used online social data to study human behavior, including modeling the interaction between social connections and geography (published in PNAS) [43] and the evolution of online friendships [45]. Other work has combined data mining and visual analysis on millions of images, including mining geographic information [44] (runner-up best paper at WWW 2009), studying geo-temporal relationships between tags [188], and producing city-scale 3D reconstructions [54, 55] (runner-up best paper at CVPR 2011). The PI has a history of successful interdisciplinary collaborations with polar science [48], biology [187], marketing [22, 23] and psychology [126, 130, 131].

The PI is also qualified for the educational and outreach components of the plan. As an Assistant Professor in the School of Informatics and Computing at Indiana University (IU), he has taught graduate computer vision and undergraduate data mining and search courses, all scoring above the 70th percentile on anonymous student evaluations, and taught three undergraduate courses as a Postdoc at Cornell. He has advised four undergraduate projects at IU, leading to several talks and papers [54, 55, 89, 130, 131] and a award [99, 100] for undergraduate research excellence, advises four graduate students, serves on ten Ph.D. student committees, and participates in various department and university outreach activities including the IU lifelong learning program [3]. He has served on program committees of 26 conferences and workshops in data mining and computer vision (including KDD, WSDM, WWW, NIPS, CVPR, ICCV, ECCV, and Multimedia).

2 Initial results

The PI's work over the last three years has begun to lay the foundation for large-scale visual social sensing, contributing to each of the above four research objectives (calibration, extraction, aggregation, application).

Mining geography. We have explored how to organize massive collections of geo-tagged photos using both visual image content and non-visual metadata including text tags and captions [44]. This work proposed data mining algorithms that automatically identified key places on earth by looking for peaks in the geotag distribution, inferred textual descriptions for each place by looking for distinctive tags, and found representative images for each place by finding similar scenes taken by many photographers. Figure 1 shows a sample result from this work: an annotated map of North America, produced completely automatically using 60 million geo-tagged Flickr photos. This work is a striking example of the power of social sensing and received attention from both the press (including Wired [163], New Scientist [24], BBC Radio, and the Guardian) and from researchers in a diverse range of fields who saw applications to their work.

Large-scale 3D reconstruction. Our work on large-scale 3D reconstruction has used large sets of images downloaded from social media sites to reconstruct detailed 3D models of a place, including an individual landmark or an entire city, without any prior information about the place [54, 55]. This work was inspired by recent success in applying structure-from-motion techniques to unstructured collections of online images [9, 79, 172]. The idea is to identify distinctive visual features that can be matched across images as likely belonging to the same scene points, and then using the constraints induced by multiple cameras viewing the

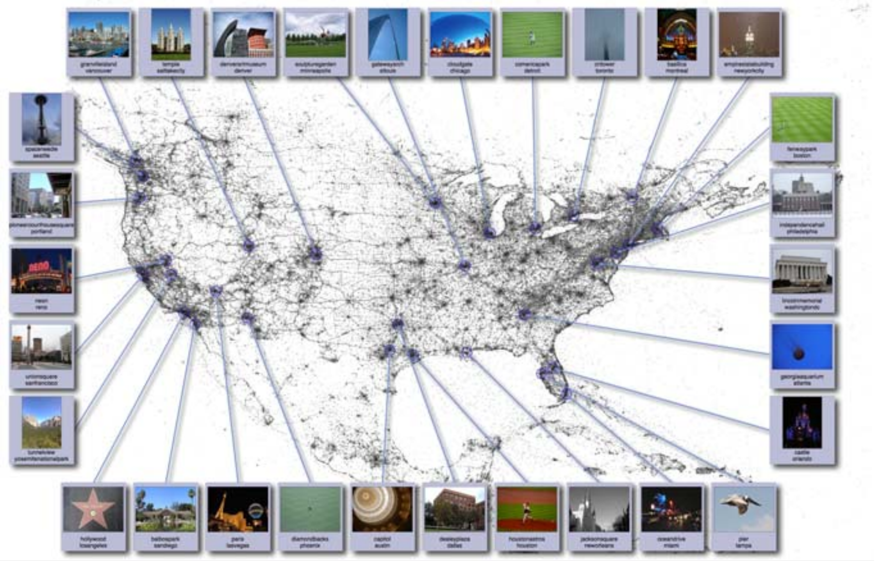

Figure 1: Automatically-generated map of North America produced from 60 million geo-tagged images [44]. For the 30 most-photographed cities, we find a text tag describing the city, a text tag describing its most popular landmark, and a representative image.

same rigid scene from different perspectives to simultaneously estimate both the 3D geometry of the scene and the 6D pose of the camera that took each image. These approaches scale poorly ($O(n^4)$ in the number of images) and often fail by finding bad local minima. Our work has presented an alternative approach that casts 3D reconstruction as an inference problem on a statistical graphical model. In particular, the model is a Markov Random Field [108] in which vertices represent camera poses and scene points, and edges represent estimated camera-camera and point-camera constraints on relative pose. The inference problem is to assign an absolute pose to each camera and point such that the relative constraints are satisfied to the greatest possible extent. While MRF inference is NP-hard in the general case, we use loopy discrete belief propagation (BP) [146] to produce good approximate solutions that then can be quickly refined using continuous optimization. Our approach is faster both asymptotically ($O(n^3)$ vs. $O(n^4)$) and in practice, notably because BP can be easily parallelized on distributed-memory clusters. Our implementation on a 200-core map-reduce cluster allows us to reconstruct scenes from tens of thousands of images, as illustrated in Figure 2.

Image understanding. Our work in extracting semantic information from images has investigated various object recognition approaches. The PI's earlier work investigated models that represent an object as a set of parts arranged in deformable spatial configurations, and introduced a family of statistical models called k-fans for which exact inference is efficient using dynamic programming, min-convolution, and branch-and-bound search [46, 47], as well as a weakly-supervised learning algorithm [49] that requires only a set of training images known to contain the object of interest. Figure 3 presents sample results of this system. Other work has focused on recognizing landmarks (including buildings, statues, and other tourist locations) in nearly 100 million online photos downloaded from social media [119]. We used structured support vector machines [176] to learn visual models from vector-quantized feature descriptors [57] for each landmark, and implemented recognition and learning in the cloud using map-reduce [182]. These approaches work well for detecting rigid, man-made objects, but new approaches are needed for highly flexible natural objects like

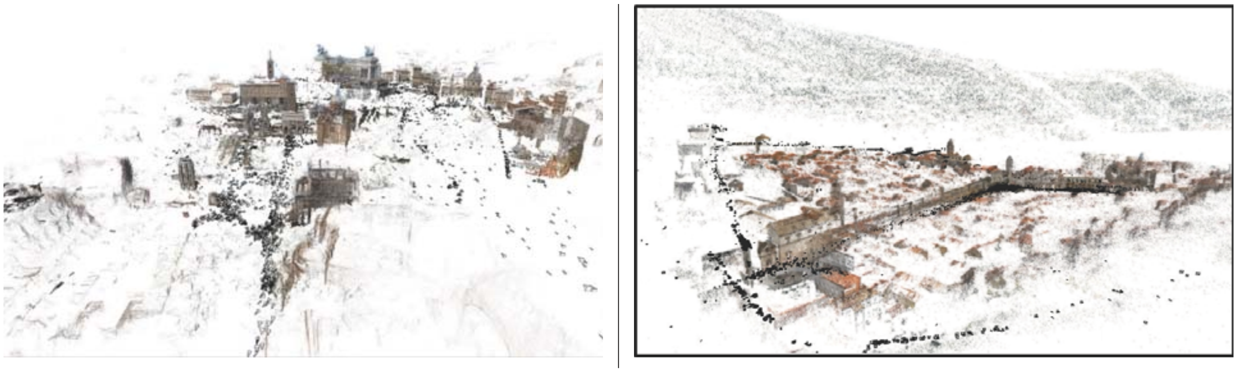

Figure 2: 3-D reconstructions of central Rome (left) and the old city of Dubrovnik (right) using our MRF-based reconstruction technique [54, 55] applied to tens of thousands of Flickr photos.

animals [69]. We are investigating the specific problem of fine-grained category recognition, like classifying between different species of birds, in which subtle differences are important and large training sets may not be available. Our work uses *attributes*, which are visual properties that are both visually distinctive but also meaningful to humans (e.g. 'yellow beak', 'red head', etc.) [70]. This approach enables interaction with people: humans can understand the learned object models, allowing object detection algorithms to 'explain' their reasoning in terms that a human can understand, while human experts can specify characteristics of new unseen classes to the system using the attributes it already understands. Figure 4 shows some results from our work in automatically learning attributes and using them to detect properties of unseen images [65].

Estimating ecological distributions from social media. We are investigating how to use social photo collections to observe properties of the natural world. This initial work has searched for evidence of snow, by using both text tags and visual analysis, in a collection of nearly 150 million time-stamped geo-tagged photos from Flickr. Using machine learning and simple Bayesian statistical tests, this work estimates the distribution of snow across time, creating daily maps of snowfall similar to those produced by a satellite [187]. Figure 5(a) shows examples of the snowfall distribution map for North America produced for three sample particular days, and compares them to satellite data from the MODIS instrument aboard NASA's Terra satellite [91]. The figure shows that these techniques have complimentary strengths: the satellite gives uniform coverage across the continent, while the Flickr analysis only works in areas with social media users; on the other hand, the satellite can only observe the ground when the sky is clear, whereas our Flickr analysis has no such restriction. A quantitative comparison of the accuracy of our snow distribution maps over a period of several years, using satellite and metereological data as ground truth, has indicated that our social sensing-based snow estimation is highly accurate and complete in cities (where the photo density is high), while the estimates are still accurate but significantly sparser in rural areas. This initial work supports the hypothesis underlying the Career plan, that mining large-scale collections of visual social media could unlock new sources of observational data about the world, with advantages that are complementary to existing observational techniques.

3 Related work by others

By its nature, the proposed project will build on and unite work in both data mining and computer vision. Here we briefly review work that relates to crowd-sourced sensing in general, as well as to the four threads of investigation that we propose to study (calibration, extraction, aggregation, application).

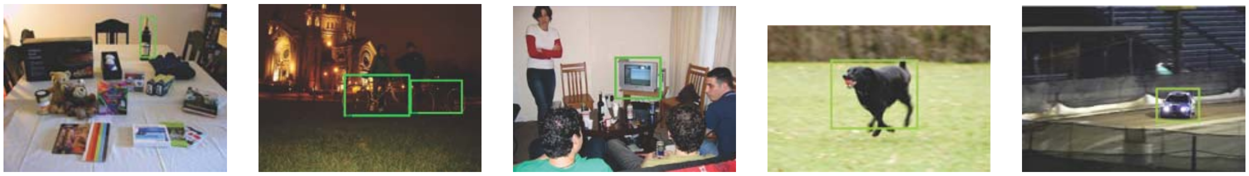

Figure 3: Results of detecting (from left) bottles, bicycles, televisions, dogs, and cars, using our object recognition system [41].

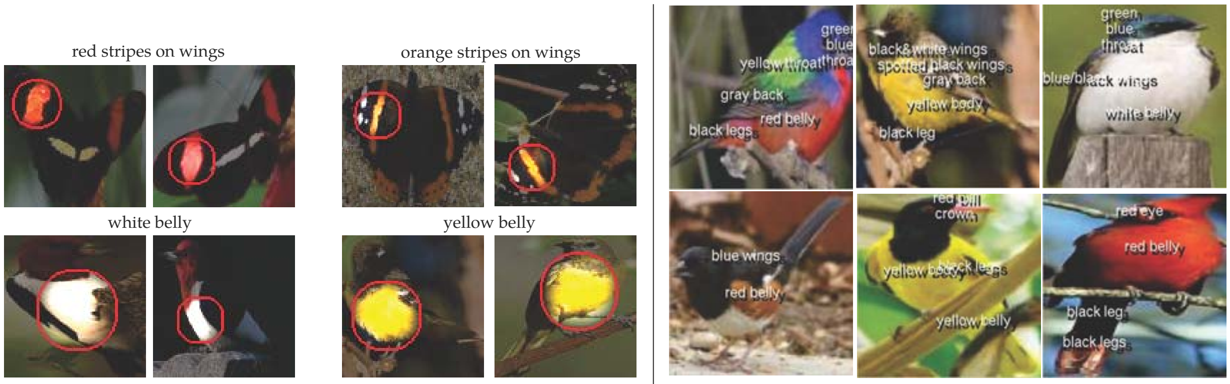

Figure 4: Results of our attribute-based recognition approach [65]. Attributes (which are visually distinctive properties that are also nameable and meaningful to humans) are discovered through automated learning with minimal human interaction (left), and then can be detected automatically in new images (right).

Mining social data. A variety of recent work has studied how to apply computational techniques to analyze online data in order to aid research in the social sciences (a young field tentatively called Computational Social Science [114]). This work includes studying how friendships form [45], how information flows through social networks [120], how people move through space [33, 43, 96], and how people influence their peers [12, 21]. Other studies have shown the power of social datasets as a source of observational data about the world itself; this work can be seen as a specific case of ‘social sensing’ [10, 157] in which observations from individual people are aggregated together. This work includes using Twitter data to measure collective emotional state [87, 139] (which, in turn, has found to be predictive of stock market moves [31]), predicting product adoption rates and political election outcomes [102], and collecting data about earthquakes [67, 160] and other natural disasters [62, 76, 161]. Others have tracked the geo-temporal distributions of social media tags, query search terms, and English words in large datasets [19, 37, 40, 83, 132, 152, 153, 158, 169, 179, 180]; we highlight two threads of work that are particularly striking examples of what can be learned by data mining huge corpora: (1) tracking search query terms related to symptoms of flu was found to accurately predict spread of the H1N1 outbreak in real-time [83] (while Twitter even predicts when a particular person will fall ill [158]), and (2) mining huge datasets of books and newspaper articles to study how society has changed over time [40, 132, 133] (called ‘Culturomics’) and to predict future events [153].

Organizing large-scale image collections. Most of the work on visual social media data has been motivated by helping users to organize large photo collections. This work includes studying the usage of text tags [85, 129, 162, 164], suggesting tags from GPS position and predicted tag co-occurrence [81, 121, 134, 135, 168], modeling relationships between tags and visual features [183], modeling and predicting relationships between users based on their images [12, 86, 116, 117, 181], and estimating geo-tags from visual image content using broad scene properties [94, 104, 165] or by recognizing specific buildings and landmarks [80, 119,

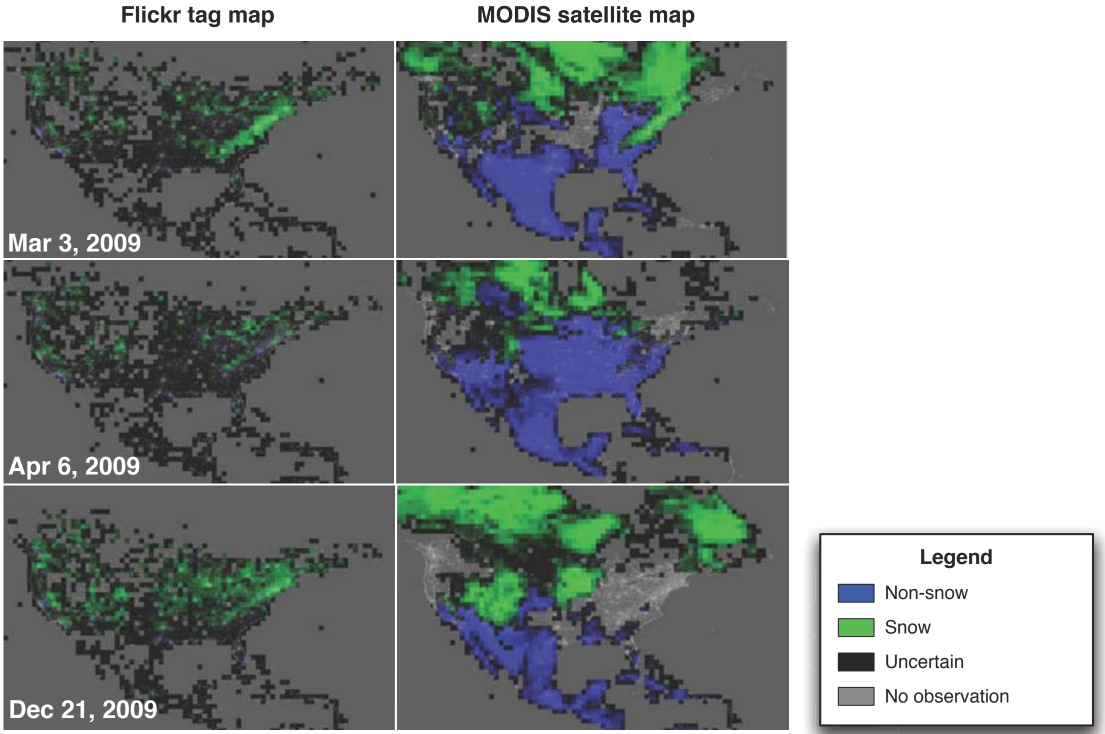

Figure 5: Estimated North American ground snow cover on three specific days using evidence from Flickr photos (left) and data from the NASA Terra satellite (right). Note that black or gray areas indicate missing data, which is usually caused by lack of photographic evidence in the case of Flickr, or by cloud cover in the case of Terra.

149, 151, 192]. Other work has tried to model and predict aesthetics and sentiments underlying social media images [73, 167]. In terms of using Flickr for crowd-sourcing, one popular problem has been to combine geo-tags, text tags, and (in some cases) image data to identify heavily-photographed places on Earth, producing rankings of popular places and automatically-generated maps [106, 155], detecting and characterizing popular events [11, 36, 66, 103, 142], or recommending tourist itineraries [16, 60, 61, 110, 141, 150]. More related to what we propose here is work that tries to classify land type using visual analysis of photos [118], work that compares images from people of different cultures to find distinctive differences [178, 185], and studies that use video data from stationary webcams to estimate weather conditions [101].

Object and scene recognition. There is a vast literature on recognition in computer vision that spans nearly 50 years; we briefly summarize the major themes here. Most modern recognition approaches use machine learning to train models from sets of labeled exemplars. At a very high level, these approaches differ in: (1) the types of low-level image features used, (2) the structure and assumptions of the object or scene models, and (3) the classifier and learning algorithm. Many popular low-level features are based on histograms of edge gradients in grayscale images, including HOG [59], SIFT [123], and SURF [25]. Scene-level features include simple color histograms [173], filter responses at different orientations and scales across the image [175], or a combination of these two [94]. Other work segments images into homogeneous regions [35, 72, 166] and then computes shape and color features [27, 90]. In terms of model types, the so-called 'bag-of-words' models, in which objects and scenes are represented as unordered sets (analogous to similar approaches in document retrieval) [58, 170] are popular, while other work imposes very coarse spatial constraints [88] or highly flexible models [46, 71, 74, 127] between features. For learning, most modern work uses discriminative classifiers like Support Vector Machines [29] or structured SVMs [176], although nearest-neighbor classifiers work surprisingly well in large datasets [30, 189]. Many variants of the above techniques exist, like creating composite scene and object models [50,95], expressing objects and scenes in terms of grammars [84, 191], or using ‘attributes’ – features that are both visually distinctive but semantically meaningful to humans [65,70,109]. Most of this work has focused on detecting man-made rigid objects like cars and buildings (perhaps because of security and surveillance applications), so performance on natural objects like plants and animals is comparatively low [69]. There have been some recent studies that have targeted natural objects, like recognizing flowers [105,138,159], leaves [26,115,184], horses [77], cats and dogs [143,156,190], fish [113], elephants [17], and a variety of other animals [8,28,32,75,111].

4 Research Plan

We propose a 5-year research plan to investigate and demonstrate the feasibility of visual observation of the world using photographs from social media. The plan builds on our existing work in this area, described in Section 2. The plan is designed around three testbed applications of visual social sensing, which will be used to validate the viability of this idea and to measure its strengths and limitations. To facilitate visual social sensing, the plan investigates three core threads: calibrating the social sensors to know where and when each observation was recorded, extracting observational information from individual photos and their metadata, and aggregating evidence from multiple photos together into accurate observations about the world.

Data. This is a data-driven project that will involve analyzing large collections of images from social media sites. As described in Section 2, the PI has extensive expertise with image analysis on large-scale datasets, including a dataset of about 150 million geo-tagged images from Flickr. This dataset was collected using the public Flickr API in accordance with the Flickr Terms of Service and IRB. The proposed project will use this existing dataset as a starting point, and we will continue to collect additional data as appropriate. If for any reason it is no longer possible to collect data from Flickr, we will switch to one of the several other photo sharing services like Panoramio; even if this is not possible, the entire proposed project can be carried out using our existing dataset or one of the other large-scale image datasets that have been made available for academic research use (e.g. [63]). Please see the attached data management plan for more details.

4.1 Validation and applications

The PI will investigate three interdisciplinary applications of visual social sensing. These are a core component of the research plan, because they will be used to verify the hypothesis that social sensing in large photo collections can create new, accurate data sources about the world that will be useful to scientists, and will identify the capabilities, challenges, and limitations of this idea. Results of these validations will help to direct the other three threads throughout the course of the project.

Studying human preferences. This study will examine the properties of photos that people upload to Flickr, in particular looking for differences across the demographic groups of the photographers (including different cultures and genders). Work in the evolutionary psychology literature has shown that in laboratory settings, people prefer different colors: women prefer redder hues, while men prefer bluer hues, and differences have also been observed across cultural lines [97, 140]. Our hypothesis, which so far has stood up in our initial work [131], is that such differences may also be observed through analysis of the millions of photos that people upload to photo sharing sites, as well as the ones they view, comment on, and ‘like.’ (See collaboration letter from Dr. TTT, Professor of Cognitive Science.)

Estimating geographic properties. This study will examine the ability of Flickr image data to predict the properties of a place, including for example geological (elevation, elevation gradient), demographic (population, population density, mean income), climate (average temperature, average rainfall), and land type (desert, forest, plain, water). We will take a data-driven approach, using machine learning to find correlations between visual and non-visual features and ground truth training data derived from GIS maps (such as the datasets of [5]). We will measure the ability of our classifiers to produce dense geospatial maps of estimates of each of these properties, evaluating their quality with respect to ground truth.

Observing ecological phenomena. This study will examine the ability of visual social sensing to estimate time-varying properties of the world that are of interest to biologists and ecologists, like the geo-temporal distribution of animals and other ecological phenomena. This will build on our initial work (see Figure 5) that considers the particular problem of estimating geo-temporal distributions of snow cover. We will consider the more difficult problems of estimating distributions of other weather phenomena like cloud cover, rain, and fog, and of particularly visually distinctive birds and flowers such as the California Poppy (which is bright orange). Here we will use satellite and herbaria data as ground truth for both learning and testing. (See collaboration letter from Dr. LLL, Professor of Biology.)

4.2 Calibrating the social sensor network

Each image on a photo-sharing site is an observation of the world at a particular point in time and space, but to make use of this observation we need accurate estimates for when and where the photo was taken. Modern cameras record photo timestamps, but these are often inaccurate because people neglect to set the clock or update it when they travel across time zones. (Note that image sharing is different in this regard from other social media systems like Twitter, where timestamps are assigned by the systems' servers and are thus guaranteed to be accurate.) Meanwhile many photos on social sharing sites have geo-tags specifying latitude-longitude coordinates, but these are also noisy: consumer GPS receivers have a precision of tens of meters under ideal conditions [20], while manually-input geotags can be arbitrarily inaccurate.

Calibrating timestamps. To calibrate timestamps of photos, we will examine *streams* of photos taken by the same user. The assumption here is that even if a camera's clock is set incorrectly, the *relative* time between photos taken by the same camera will be correct. Thus calibrating the timestamps involves aligning photo streams to a global time coordinate system by shifting them by an unknown offset. To estimate this offset, we propose to build joint models of geolocation, time, tags, and visual content, so that we can evaluate whether the visual content of a given image agrees with the degree of sunlight expected at that place at that time (given the expected sunrise and sunset times). While this estimation will be noisy for any particular estimate, our hypothesis is that for sequences of several images the alignment can be estimated accurately.

Calibrating geo-tags. To estimate accurate geo-tags, we will leverage the PI's work on large-scale 3D reconstruction (Section 2). A side effect of 3D reconstruction is that a very accurate camera pose for each image is estimated, with typical geolocation errors in the centimeters [55], and new images can be geo-located very accurately by aligning to an existing 3D model. Since it is not (yet) practical to build 3D models of the entire world, we propose to reconstruct models for the 1,000 most photographed landmarks on Earth (which we have mined and studied in [44]), and use these models to very precisely geo-locate *some* images. These images will act as high-confidence 'anchors', weighted heavily during social sensing, while evidence from images that cannot be precisely geo-located will still be incorporated but with lesser weight.

Larger-scale models. To reach our eventual goal of precisely geo-locating *all* images, we need more scalable techniques for building the 3D models. Our work on large-scale reconstruction has significantly advanced the state of the art by proposing an efficient framework using graphical model inference, but our approach is limited by the use of discrete state spaces to encode latent variables: as the physical space to be reconstructed grows, the size of the discrete messages exchanged during belief propagation (BP) [146] become unwieldy, requiring too much memory and computation time. We are investigating alternative parameterizations that will allow 3D reconstruction to scale to much larger image sets, including using mixtures of von Mises distributions (which are analogous to normal distributions on a k-sphere) for parameterizing the rotational portion of pose, and mixtures of normal distributions for parameterizing positions. The use of mixture models is important because of the large degree of noise in this problem, which requires techniques that are robust to outliers [55], but how to use them efficiently with BP is an open question that we will study.

4.3 Extracting semantic information from images

Tags and metadata. Captions, comments, and image text tags present very convenient (albeit limited) means of extracting semantics from images, and we plan to use them during early phases of the project. As shown in Section 2, simply plotting distributions of a particular tag (like 'snow') across space and time yields some meaningful signal. We will investigate more principled approaches like learning correlations between tag presence and the properties we wish to observe (such as weather or geographic properties of a place).

Scene classification. We plan to investigate techniques including color histograms [173] and GIST [175] to extract scene-level features from images, again learning classifiers using these features to predict properties of interest to our applications. We will employ two types of training and test data. As in our past work [119], we can cheaply create large-scale labeled training data by using the geo-tags on photos to look up ground-truth attributes in GIS maps (e.g. whether a photo was taken at a place with sand, forest, or snow, etc). This ground truth data will be noisy, due both to errors in geo-tags and also ambiguities (like photos taken at a snowy place but indoors). Our second approach will be to use hand-labeled training and test data, produced by undergraduate assistants and/or Mechanical Turk tasks, which will be more accurate but smaller in scale.

Object classification. For applications like producing geo-temporal distributions of ecological phenomena, we will need techniques for extracting object presence information from images. Here we will use our recent work on attribute-based recognition (Section 2), which finds image features that are both visually distinctive but also semantically meaningful to humans. These approaches have been found to work better than traditional techniques in detecting natural objects like flowers and animals, and we plan to further characterize and improve upon their performance in this task.

4.4 Aggregating evidence from multiple photos

After evidence has been extracted from individual images, techniques are needed to aggregate this evidence into coherent estimates (e.g. whether there is snow on the ground at a particular place). These techniques need to be robust to both noisy observations and various sources of bias.

Characterizing the bias. We will first investigate and attempt to quantify the various sources of bias. We believe bias will arise from several sources, including: (1) geographical bias caused by uneven population distribution across the globe, (2) observer bias caused by the fact that social media users are younger and more tech-savvy than the general population, and (3) bias towards photographing certain events and objects more than others, such as ignoring common objects in favor of the unusual. We plan to quantify these sources of bias by using our initial snow cover detection work (Section 2), exploring for example whether more photos are taken of unusually-timed snow storms, and the extent to which this varies across the demographics of the photographer. (Public Flickr profiles include some demographic information like gender and location; for studying the effects of age we will contact a subset of high-activity users to ask their broad age group.)

Robust aggregation. We will then investigate techniques for robust aggregation of evidence into accurate observations. We will explore statistical approaches that assume that evidence from different users is independent. For estimating classes, like land cover type, we will use simple voting, while for estimating presence or absence of an object, we will use a Bayesian framework to calculate odds ratios given the quantity of evidence with respect to a prior. We will investigate using machine learning techniques to correct for the biases discovered above, by weighting evidence depending on its estimated accuracy and potential bias.

5 Broader Impact: Education, Outreach, and Collaborations

The proposed project unites computer vision, applied machine learning, and large-scale data mining with interdisciplinary applications, creating unique opportunities for outreach and education. The highly-visual nature of the work appeals to undergraduate students and general audiences and makes it easy to explain highly technically concepts at an intuitive level. (As evidence of its ability to spark the imagination of general audiences, several of the PI's recent papers have garnered attention of the popular science press and blogs including New Scientist [18, 24, 107], Wired [163], and Slashdot [7].) The data-intensive nature of the project aligns well with NSF's Big Data priority [137], and will help to train undergraduate and graduate students for the rapidly-expanding job market for data-intensive computing. Interdisciplinary collaborations will magnify the impact of the project across a variety of disciplines inside and outside of Computer Science.

5.1 Education and outreach goals

The proposed Career plan will focus on four specific education objectives:

1. Increasing undergraduate involvement and interest in research. Exposure to research is a critical component of undergraduate education: research helps to spark student interest in advanced computer science topics, while teaching both technical and life skills about teamwork, independence, and time management [112, 122]. Moreover, research experiences help to expose students to the opportunities available in graduate school and to identify talented students interested in research trajectories, which in turn helps to ensure a strong pipeline of academics to feed the long-term scientific educational and technical needs of the nation. (The PI feels a personal devotion to this goal because his involvement in a research project as an undergraduate was the turning point that launched his academic career.) Unfortunately, the PI's department at Indiana University (IU) does not have a strong history of promoting undergraduate research, and although the school is actively working to change this, fewer than 10% of our undergraduate students go to graduate school and fewer than 2% go on to study Computer Science [34].

2. Preparing students for today's statistical and data-intensive jobs. Experience with statistics, machine learning, and data intensive computation are becoming important prerequisites for careers in a range of disciplines, due in part to the so-called Big Data Revolution [13, 137, 147]. The curriculum at IU (and presumably at many other universities) has not kept pace with these recent developments in computer science; for example, neither our undergraduate nor graduate CS curricula require any courses in statistics, probability, data mining, or machine learning, and current course offerings in these areas are slim [4].

3. Attracting female, minority, and other under-represented groups to computing. Despite being Indiana's second-largest racial group at about 9% of the population, African-Americans account for just 2% of students in the graduate computer science program at IU [177]. Meanwhile, just 16% of students in our program are women. The PI feels that it is simply not acceptable for the CS program in our state's flagship state university to be so poorly representative of our population. This lack of diversity puts IU and the greater computing community at a disadvantage, starving us of important insights and perspectives that could be brought to bear on our field.

4. Educating general audiences, particularly older citizens, about computing. A so-called "grey digital divide" is growing between young people who use technology regularly, and many senior citizens (particularly those in rural areas) who are ill-informed and distrustful of online technology including social media and data mining [38]. This is particularly unfortunate because recent evidence suggests that parents are a key factor in high schoolers' interest in science and technology careers [92].

5.2 Ongoing and planned education activities

The PI has worked towards the above four goals throughout his first two years at IU. This work will be expanded in the 5-year Career plan through several education and outreach activities.

Undergraduate teaching and curriculum development (goals 1, 2, and 3). The PI has taught two iterations of Info I427, a senior-level information retrieval and data mining course designed for students in both our Computer Science program and our Informatics program (which is IU parlance for 'information science'). This is a hands-on course culminating in a final project that requires students to implement all aspects of a web search engine, including the web crawler, indexer, retrieval with TF-IDF, ranking with PageRank, and a web interface. In addition to these technical components, the class also examines the ethical, legal, and financial aspects of web search. This course has received consistently high reviews in anonymous student evaluations (3.5 out of 4.0 on the quality of course and instructor), with one student calling it "the best course so far at IU – a course finally with content and techniques that are current and relevant." One student, MMAA, extended his course project to create a small start-up company that aggregates and ana-lyzes activity of Saudi users on Twitter [6], and has grown successful enough to attract media attention [14]. The PI has found that existing course materials are not adequate for this type of interdisciplinary class: the standard information retrieval and search textbooks (e.g. [56, 128]) are very good but are too technical for this interdisciplinary student audience, and do not cover non-technical topics of search like business models and ethical issues. The PI has been writing lecture notes and other materials for this class, and plans to release them on a public webpage for others to use.

The PI will also develop a new senior-level undergraduate Computer Science course on computer vision. The course will be loosely based on his successful graduate vision course offering, but will emphasize core techniques like statistical models and machine learning that can be applied to many different fields, using computer vision as a motivating application.

Finally, the PI is part of a three-person faculty committee to design a new Data and Search Specialization within the Computer Science major. This new concentration will complement the existing more traditional concentrations (theory, programming languages, artificial intelligence, and systems) to prepare students for the new era of 'big data'-related careers. In addition to designing the specialization, part of the committee's mission is to educate incoming and first-year students about data mining and search, making them aware of the area and the opportunities in it. The PI has so far given four such presentations in introductory CS courses and orientation events, and will continue this outreach throughout the 5-year plan.

Graduate teaching and course development (goal 2). The PI teaches CS 657, a graduate-level introductory course in computer vision. This course had not been recently offered when the PI arrived at IU in 2010, so the PI redesigned it to reflect modern progress and challenges (such as topics in Markov Random Fields and other graphical models, large-scale structure from motion, part-based object models, etc). Anonymous student reviews have been in the 90th percentile in the university, with comments including "The course is very modern and the material explained well," "This class was a wonderful experience and the professor was very good in all aspects," "The projects were really very interesting and challenging," and "The course is very valuable for my research." One of the course projects on mining in large-scale image data eventually resulted in a conference paper at WSDM [188]. The PI plans to continue teaching this course biennially.

The PI is currently developing a new graduate-level course on advanced machine learning, with a particular emphasis on structured learning and probabilistic graphical models. Despite the increasing ubiquity of machine learning and statistical inference techniques in both research and industry, IU does not currently offer a course that covers these topics, placing graduates of our program at a significant competitive disadvantage. The course will use the textbooks of Koller [108] and Bishop [29], and course projects will integrate with the large-scale visual mining project of the proposed project.

Undergraduate research experiences (goals 1 and 3). The PI has made undergraduate research a priority and has supervised or co-supervised research with four undergraduate students during his two years at IU. Of these projects, one resulted in a poster presentation at an undergraduate research symposium [89], one resulted in a talk [130] and poster [131] at a psychology conference, and one contributed towards a paper at a top vision conference [55] and a journal article [54]. The fourth project, with undergraduate AAA , won an IU Provost's Award for Undergraduate Research [99]; AAA has formed a start-up company to commercialize the work and has found seed funding [100]. The PI also supervised 7 undergraduate projects as a Postdoc and PhD student at Cornell, with these students later joining top institutions like Washington, MIT, Stanford, and Lincoln Labs. To continue this work, the PI has budgeted to hire 2 undergraduate students per year to assist in the proposed project. The PI will initially involve these students in very concrete tasks, such as web programming for the project websites, and then gradually move them into more independent and research-oriented roles as their interests and abilities allow. The PI will also pursue REU funding for additional undergraduate student involvement during the summer.

Outreach to HBCUs (goals 1 and 3). The PI plans to make two visits to HBCU universities, during years 1 and 3, in order to give talks and recruit students both to participate in the PI's summer REU program and to encourage them to apply for graduate school. Dr. XXX , PI of the Alliance for the Advancement of African-American Researchers in Computing (A4RC) [1], has agreed to facilitate these visits and to help the PI recruit students into his research program (see collaboration letter from Dr. XXX).

Outreach to women (goals 1 and 3). The PI's college at Indiana University has met the aggressive goal of doubling female undergraduate enrollments in the span of 18 months [98], as part of the National Center for Women and Information Technology's PaceSetters program [136] through aggressive recruitment of prospective students and support for those in the program. The school has set a new goal of doubling enrollments again. To help achieve this goal and to help increase female enrollment in our graduate program, the PI will recruit promising female students to participate in the undergraduate research program, with the help of Assistant Dean of Diversity BBMM (see collaboration letter).

Outreach to senior citizens (goal 4). To help bridge the so-called 'grey divide' separating older and younger generations in their use of social media and online technology, the PI will give annual workshops at IU's Mini University outreach program [3], which invites alumni and community members to campus for a week each summer to attend lectures from IU faculty volunteers. The PI will also give an annual workshop at a local retirement home, Meadowood Retirement Community, for the center's Issues and Experts luncheon lecture series. (See letter from MMJJ and WWBBB , Directors of IU Lifelong Learning.)

5.3 Evaluation of educational and outreach activities

The PI will measure progress on his educational and outreach goals by conducting an annual quantitative evaluation. To assure that this evaluation is conducted in a scientifically rigorous and unbiased manner, the PI will subcontract the Indiana University Center for Evaluation and Education Policy (CEEP) [2], a leading national center on education assessment, to design an evaluation plan and protocol based on surveying participants in the proposed education and outreach activities (see collaboration letter from Dr. MMM, Associate Director of CEEP). Funds have been budgeted during year 1 for faculty researchers from CEEP to help design the initial protocol, and then again during year 3 to provide feedback from the first two years of the Career plan and to make adjustments to the evaluation framework as needed.

The PI will conduct yearly evaluations to answer questions including: (1) Are the graduate RAs graduating and/or making progress towards their degree? (2) What percentage of undergraduate research participants apply to graduate school and what fraction are admitted? (3) What percentage of undergraduates in the PI's courses use the taught concepts in job interviews, what fraction of graduated students use the concepts in their jobs, and what topics should be added or removed? (4) How many minority students participate in the research program, how many of the undergraduate minority students apply to and are admitted to graduate school, and how many graduated minority students are employed in a computing field? (5) Do participants in the outreach workshops leave with a better understanding of data mining and related technologies, and how could this be improved? (6) What do undergraduate and graduate course evaluations indicate about the courses and how they can be improved? To enable this evaluation, the PI will track contact information for a sample of each relevant population, again using methodology recommended by CEEP.

5.4 Planned collaborative activities

Broader impacts of the proposal are highlighted by the collaborations that will take place.

Computational frameworks for large-scale data mining. The PI has an ongoing collaboration with Dr. QQQ, an expert on large-scale data-intensive distributed computational frameworks, including grid and cloud computing. Her group is developing the Twister system [68, 78], a variant of traditional map-reduce that is optimized for iterative jobs (by smart scheduling of map-reduce tasks across algorithm iterations and by caching intermediate data in (distributed) system memory). Dr. QQQ has agreed to lend her systems expertise to our project, including giving us access to large-scale compute resources including the NSF Teragrid and XSEDE, in exchange for us providing her with realistic, large-scale data analysis problems with which she can test her computational frameworks. (See collaboration letter from Dr. QQQ).

Studying human behavior and preferences. With Dr. TTT, Professor of Cognitive Science and Director of the aaa Laboratory, the PI is leading an interdisciplinary team of undergraduate and graduate students to use photos from social photo sharing sites as a means to study human behavior at global scales. This collaboration has already resulted in an oral talk [130] and poster [131] at an international psychology conference. (See collaboration letter from Dr. TTT.)

Mining ecological data. The PI has an ongoing collaboration with Dr. LLL, Professor of Biology at sss University and director of the ggg Project, a leading citizen-science initiative. The goal of this collaboration is to use images and metadata from social photo sharing sites like Flickr to reconstruct information about the environment, such as where and when flowers are blooming or where there is snow on the ground. This collaboration has led to a recent paper at WWW 2012 [187]. The broader impact of this collaboration could be a new source of data for studying ecology and phenology by leveraging the vast number of online photos. (See collaboration letter from Dr. LLL.)

5.5 Outreach to the research community

In addition to publishing our research in journals and conferences, we will release software, benchmark problems, and data developed in this research on the World Wide Web. In particular, as part of the collabo-ration with Prof. QQQ, the PI will release a software toolkit for large-scale image analysis using map-reduce, and will offer a tutorial on these computational frameworks at a vision conference during Year 4. To spark engagement between the computer vision and data mining communities, the PI will also organize two work-shops on visual data mining, one at a computer vision conference (such as CVPR or ICCV) during Year 3 and one at a data mining conference (such as WWW or KDD) during Year 5.

6 Timeline and summary

A summary of the PI's 5-year Career plan is presented in Table 1. While the table lists research, outreach and education activities separately for clarity, the three components will be tightly interconnected: research results and challenges will be integrated into education, outreach will encourage others to study visual social sensing to accelerate research progress, and graduate and undergraduate students educated under the plan will provide the energy and skills for the research in this project and beyond.

Table 1: Summary of the five-year Career plan.

	Research	**Education**	**Outreach**
Year 1	– Build 3D models – Tag extraction – Color preferences – Study bias	– Develop grad ML – Teach grad CV – Teach ugrad DM – Evaluation	– Summer REU – Lifelong workshop – Visit and recruit at HBCU – Large-scale vision WWW resources page – Evaluation
Year 2	– Estimate geotags with 3D models – Scene classification – Bayesian aggregation – Geospatial property estimation	– Launch ugrad DM course resources WWW page – Teach grad ML – Teach ugrad DM – Evaluation	– Summer REU – Lifelong workshop – Evaluation
Year 3	– Timestamp alignment – Scene classification (cont'd) – ML aggregation – Weather estimation	– Teach ugrad DM – Teach grad CV – Develop ugrad CV – Evaluation	– Summer REU – Lifelong workshop – Workshop at CV conference – Visit and recruit at HBCU – Evaluation
Year 4	– 3D reconstruction parameterization – Attribute recognition – Flowering distribution estimation	– Teach ugrad DM – Teach grad ML – Teach ugrad CV – Evaluation	– Summer REU – Lifelong workshop – Large-scale vision tutorial at CV conference – Evaluation
Year 5	– 3D reconstruction parameterization (cont'd) – Attribute recognition (cont'd) – Animal distribution estimation	– Teach ugrad DM – Teach grad CV – Teach ugrad CV – Evaluation	– Summer REU – Lifelong workshop – Workshop at DM conference – Evaluation

(Legend: CV=computer vision, ML=machine learning, DM=data mining and web search, REU=Research Experience for Undergraduates, HBCU=Historically Black College or University.)

7 Results From Prior NSF Support

PI was supported by an NSF Graduate Research Fellowship (August 2003–August 2006), which funded his Ph.D. at Cornell and supported his thesis work on object recognition [41,46,47,49,50,52].