SIGMA: Scalable Image Graph Matching and Analysis: A top-down framework for image classification, localization, and 3-D geo-registration

Karl Ni, Scott Sawyer, Alex Vasile, and Nadya Bliss
MIT-Lincoln Laboratory
{ karl.ni , scott.sawyer, alex.vasile , nt } @ ll.mit.edu

Introduction

Because any individual digital photograph may contain considerable amounts of information, the ability to understand and extract scene information is advantageous to many communities, including, but not limited to, online social networking sites, the intelligence agencies, and logistics and analytics corporations dealing with large-scale data mining. One category of scene information is geo-spatial data that can provide the location of people, objects, and the originating camera. Sometimes, meta-data (including GPS coordinates or annotated landmarks) that provides such scene information of digital photos, is, either intentionally or unintentionally, excluded, unavailable, or not yet ascertained. In such cases, it is up to the human to manually annotate the photograph, which can be tedious and often impossible given the throughput of a system.

In the absence of meta-data, depending on the image content, how many, how salient, and which features are available, and the disparity between training and testing, determining a location that corresponds with a scene may occur at various refinement levels. It is logical to establish a hierarchy of what can and cannot be done. Starting at the coarsest level to the finest level, an image can be related to a set of geo-coordinates with some degree of confidence. For a given test image, the set can be a large collection of locations, each of which has a lower probability of being correct. Or, the set can consist of a single location that is extremely likely to be both precise and accurate.

Thus, it makes sense to create a framework that assesses an image at several degrees of localization potential. For images that may not necessarily provide enough features to be discernable from several different locations, we can provide candidate locations that have a higher probability being co-located. For other images that have distinct landmarks that are spatially unique, it is likely that they would match one-to-one to singular geo-coordinates.

Establishing such a framework also makes sense from a computational point of view. Because several models relating to specific locations can exist, comparing over the vast space of all possible images can be infeasible. Logically, paring down the search space using coarse geo-location models with rough spatial descriptors would make the problem much more tractable.

The framework involves complicated training and setup procedures in order to promote real-time exploitation. Obviously, for generalization purposes, the required data set is an extensive collection of tens of thousands of images, with location information, at high-resolution. To support the training operations, resources are distributed across several compute nodes on Lincoln Laboratory's GRID cloud computing infrastructure.

The proposed approach relies on several advances in computer vision that have been made over the past ten years. Specifically, image classification and registration techniques form the foundation of our system architecture. We modify and improve upon conventional approaches in both problem spaces, while fitting them to a proposed framework. This framework is presented with results that show our program capability that identifies images and where they might belong in the world.

Top-Down Framework

The three basic levels of geo-registration are depicted in Table 1. Each of the levels utilizes three stages in processing, shown in Fig. 1. The first takes an input image, and extracts relevant features from it. The second stage matches the extracted features against a pre-collected database. Finally, the third stage assigns an absolute geo-coordinate or set of geo-coordinates to the matched features, and by extension, to the pixels of the input image.

Table 1: Scene Categorization

Level	Result	Methodology
Coarse	Candidate Locations & Confidence	Probabilistic Classification
Medium	Error radius around Geo-coordinates	Matching Localization
Fine	Positioning and pointing direction	3D Georegistration & Pose Estimation

Figure 1: Stages in Geo-registering an Image

In geo-registering a single input image, it is simplest to conduct a top-down search from coarse classification to fine geo-registration. So, the first step, coarse classification, is defined as finding the set of potential geo-locations that with distributions similar to the input image. Of course, a training set of images at large-scale locations must be present in order to build the comparison distributions beforehand. Each of the set of locations will have an

This work is sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

associated spatial radius, a prior probability of occurrence, and a posterior probability that includes the input image. This will enable the medium localization level to find images that the input is similar to. If there are enough of these images, then geo-registration at the fine level is possible.

Coarse Classification

Empirical studies have shown that holistic processing of an image yields a better understanding of the objects within a scene because of the induced context that surround the objects. The perception is that knowing a lot about one thing is not as good as knowing a little bit about everything. The implication in our problem is that training a detector to learn a specific landmark will induce false alarms while knowledge-based scene classifiers operate much better.

Figure 2: Original image of a bedroom class

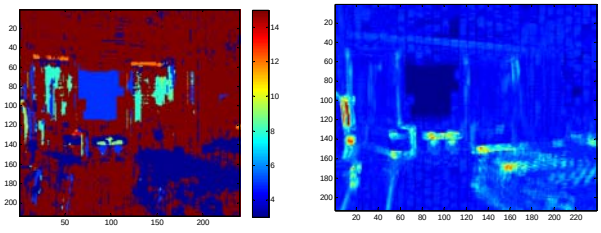

Figure 3: Bayes classification and confidence

Coarse geo-registration begins by using a semi-supervised methodology that associates labels to images. There are several databases [3] with annotated images relating to the image's content. Each image may contain multiple annotations per images, meaning that not all pixels in a single image relate to one concept. Using probabilistic distribution modeling and Bayesian decision theory, specific locations can inherit a certain pixel distribution and individual pixels can be related to concepts.

Fig. 2 and Fig. 3 provide an accurate segmentation, without having to manually segment the image. The left figure shows the actual classification of items within a scene, while the right denotes the confidence of the classification. Using the distribution of classified objects with associated semantic concepts, it is possible to build a metric that recognizes the overall geo-location. So in Fig. 2 and 3 for example, for intra-image segmented semantics of "CURTAINS", "BED", "PLAID", "CHAIR", and "WALLPAPER" on the left of Fig. 3, "BEDROOM" was extracted has a highly probable location.

Mid-Localization

Fig. 1 describes the second step after extracting relevant features as matching features to a database that is descriptive of the spatial location. This is a simple procedure, but the implications are that pictures can be roughly placed spatially provided that the training set to which they are compared are, themselves, understood in a geo-spatial sense. This becomes clear in Fig. 4, where images have been geo-registered, and an input image matches features to the remainder of the graph. Previous works [1, 2] describe this matching procedure in detail.

Figure 4: Matched image to a geo-registered image graph

3-D Geo-Registration

It is possible to build geometry, given that one knows the 3-D features that extracted features from an input image relate to. Therefore, the final stage in the coarse to fine framework is the 3-D pose estimation of an image in a registered 3-D space. Using correspondences, it is possible to build a projection matrix that describes several parameters that include scale, translation, rotation, and camera intrinsic estimations.

Figure 5: Geo-registered targets and cameras

References

We would like to acknowledge several people that were instrumental in the success of this program. Noah Snavely has made our code available to us. Peter Cho has provided several visualization tools and insights. Luke Skelly has generated several ideas and aided in the algorithmic study.

[1] Z. Sun, N. Bliss, & K. Ni, "A 3-D Feature Model for Image Matching", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2009

[2] Y. Li, N. Snavely, and D. Huttenlocher, "Location Recognition using Prioritized Feature Matching", *Proceedings of the 15th European Conference on Computer Vision* (ECCV), 2010

[3] P. Duygulu, K. Barnard, N. d. Freitas, and D. Forsyth "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", *Proceedings of the Seventh European Conference on Computer Vision* (ECCV), 2002