Hierarchical Sparse Coding With Geometric Prior For Visual Geo-location

Raghuraman Gopalan
Dept. of Video and Multimedia Technologies Research, AT&T Labs-Research, Middletown NJ USA.

Inferring which location co-ordinates an image corresponds to is a challenging problem in visual recognition, which has applications in areas such as surveillance, geomapping, landform modeling among others. It differs from other recognition problems pertaining to, say objects, scene, and faces, in that it is not enough to model visual appearance alone since two images that have similar appearance can correspond to different locations, and two images having dissimilar appearance can come from neighboring locations. While initial efforts investigating this problem began atleast two decades ago [3], substantial progress was seen only in the last decade due to the availability of large quantity of data and the related progress in data-driven techniques [4]. Most of these methods assume hand-crafted features and pursue statistical learning to discriminate images across locations.

Our approach addresses this problem using principles from automated representation learning, and we pursue this objective with both geometric and statistical flavors to learn transformational priors that help infer novel locations. (i) Firstly we understand how the appearance information of training images correlate with their location information. To this end we cluster the images based on their appearance features to form the appearance space and cluster the images based on their location information to obtain the location space. We then bridge the two spaces by considering the subspace spanned by each cluster and then employing the notion of parallel transport on Grassmann manifold to produce a collection of possible transformations that map information contained in subspaces from one space on to another. (ii) We use this geometric information as a prior for initializing a hierarchical sparse coding scheme, which automatically learns representations that compactly represent each image from the large pool of information conveyed by the geometric prior. We then train discriminative classifiers (SVM) on the sparse codes to infer location information of the test image. An overview of our approach is given in Figure 1.

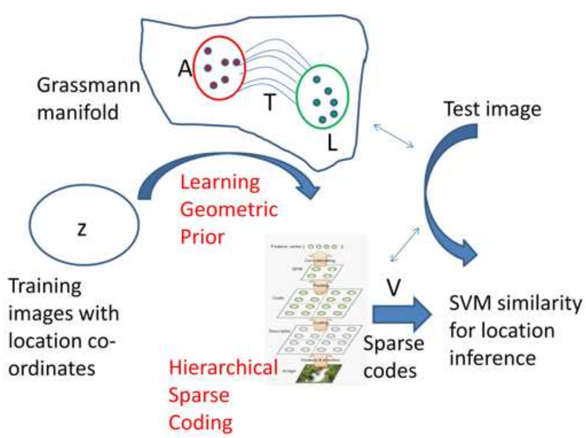

Figure 1: Given training data z with location labels, we cluster the data to form appearance space $\mathcal{A}$ and location space $\mathcal{L}$, and derive geometric transformations between them using parallel transport $\mathcal{T}$ on the Grassmann manifold. Shown on the manifold are points corresponding to subspaces obtained from each cluster. We then learn hierarchical sparse codes by initializing it with the geometric prior and perform location estimation of test data using SVM similarities learnt on sparse codes of the training data.

We believe an explicit modeling of transformations across location and appearance information is essential to address the challenging problem of location recognition. While many existing approaches consider this problem as just another classification problem, we empirically demonstrate that our geometric prior is beneficial to such techniques as well. Since such a prior increases the amount data to be processed and the dependency between them by several folds, our feature learning solution using hierarchical sparse coding brings in efficiency with regards to these challenges. We tested our approach on the publicly available earth-scale im2gps dataset [2] and more focused city-scale San Francisco dataset [1] and obtained an improvement in geo-location accuracy of upto 20% over previous best methods. Some results on the test data are shown shown in Figure 2. The construction of our approach also lends itself to integrate heterogeneous data modalities such as geotags and videos, in addition to images, and we demonstrate its utility with experiments on the MediaEval dataset. Detailed results on these datasets, as well as analysis on the robustness of the approach to parameter tuning and on the utility of design choices pertaining to the geometric prior and hierarchical feature learning are provided in the main paper.

Figure 2: Examples of test data with correct location estimates from our approach. Many images do not come from common tourist places.

We also study a relatively under-addressed problem of transferring knowledge gained on recognizing certain locations to perform inference on never-seen-before locations. This was made feasible as our model is generative and thus captures holistic trends on how appearance information flows across locations, and we could utilize it to cluster novel locations to a reasonable accuracy. Such a scheme can also serve as a means to provide weak supervision on unlabeled data. Through this work we stress that while deep statistical feature learning methods are very efficient in handling big data problems, it is equally important to feed them with strong geometric priors pertaining to the actual application that one is trying to address.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] David M Chen, Georges Baatz, K Koser, Sam S Tsai, Ramakrishna Vedantham, Timo Pylvanainen, Kimmo Roimela, Xin Chen, Jeff Bach, and Marc Pollefeys. City-scale landmark identification on mobile devices. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 737–744, 2011.

[2] James Hays and Alexei A Efros. Im2gps: estimating geographic information from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[3] Hatem Nasr and Bir Bhanu. Landmark recognition for autonomous mobile robots. In *International Conference on Robotics and Automation*, pages 1218–1223, 1988.

[4] Georg Schroth, Robert Huitl, David Chen, Mohammad Abu-Alqumsan, Anas Al-Nuaimi, and Eckehard Steinbach. Mobile visual location recognition. *IEEE Signal Processing Magazine*, 28(4):77–89, 2011.