SpiritTagger: A Geo-Aware Tag Suggestion Tool Mined from Flickr

Emily Moxley
Vision Research Lab
University of California,
Santa Barbara
emoxley@ece.ucsb.edu

Jim Kleban
Vision Research Lab
University of California,
Santa Barbara
kleban@ece.ucsb.edu

B.S. Manjunath
Vision Research Lab
University of California,
Santa Barbara
manj@ece.ucsb.edu

ABSTRACT

Geographically referenced, or "geo-tagged," photo data sets offer tantalizing potential for automated knowledge discovery in the world. By combining tag reranking based on geographic context with content-based image analysis we are able to suggest geographically relevant tags for photos newly tagged with GPS coordinates. These tag suggestions could be used to help users organize their photo collections or improve retrieval systems. Our algorithm weights labels that correspond to pertinent objects, events, neighborhoods, and activities in a region. While previous work with geo-tagged images has focused on representative views of landmarks or estimating location, our tag suggestion tool, SpiritTagger, suggests tags that reveal an insight into the spirit, or *genius loci*, of a city or region. Experiments on a data set consisting of over 100,000 Flickr photos in Los Angeles and Southern California show that our geographically relevant tag suggestion tool provides a significant improvement in precision-recall performance over baseline image-based similarity suggestion.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.2.10 [**Artificial Intelligence**]: Learning; H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Algorithms, Experimentation, Human Factors.

Keywords

image annotation, data mining, geotagging, photo collections, social media

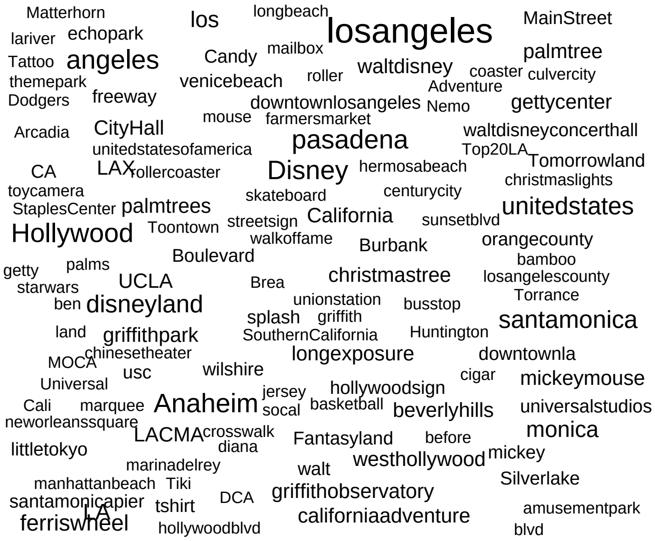


Figure 1: Cloud of prevalent tags extracted for the urban region of Los Angeles. Font size is proportional to the learned tag importance in the region. Important tags are not limited to place names. There are also relevant scene categories like "freeway" and "crosswalk" and objects like "skateboard" and "rollercoaster."

1. INTRODUCTION

Advancements in affordable cameras, bandwidth, and storage have allowed digital photo sharing to boom. Much of the photo sharing has been done at one of a few online repositories such as Panoramio[12], Flickr[2], or Webshots[17]. Our work attempts to provide understanding about urban and regional locations through the utilization of community annotations. Research into online media communities explores learning that can be done using the annotations generated by the users. Shirky[14] outlined the importance of annotation research derived from online media websites as it allows for a dynamic, evolving understanding of the world.

Interesting applications have emerged which also utilize annotations in image data with accompanying world coordinate information. Hays and Efros[3] attempt to predict the latitude and longitude of new photos by clustering nearest neighbor results in a database of over 6 million Flickr photos. More related to our work, Naaman *et al.*[8] developed a system that suggests a geographic or text annotation by searching a communal database. The suggestion process is driven by text or geo-reference only. Kennedy *et al.*[5]

MIR'08, October 30–31, 2008, Vancouver, British Columbia, Canada.

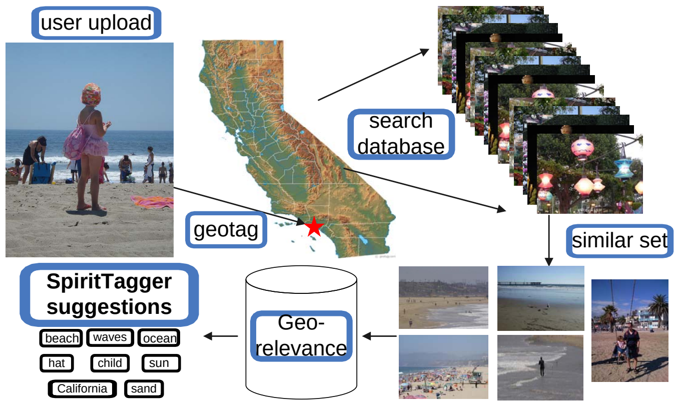


Figure 2: System Diagram for SpiritTagger. System uses geo-tagged photo to find similar images in the geographic area. Then the annotations of the similar set are mined and reranked according to georelevance. SpiritTagger takes top annotations and gives them to the user as tag suggestions for their uploaded photo.

use community annotations and clustering to generate geographic (e.g., landmarks) and temporal (e.g., events) labels. These labels are employed in World Explorer[1] for visualization. However, the idea of extending annotations to new images based on representative location knowledge and image similarity is not explored.

Our tool, SpiritTagger, suggests tags which capture the spirit of a location. For illustration, Figure 1 shows a cloud of representative tags for Los Angeles with font size proportional to local importance. While neighborhood analysis or a geographic information system could discover tags like "Santa Monica," they would not necessarily suggest tags such as "sunglasses" or "shopping" that, given the types of photos taken in a region and corresponding visual features, are likely present in a photo. SpiritTagger discovers such tags by comparing local prevalence of tags to global distribution.

A system diagram is found in Figure 2. The process begins when a user adds a new image along with an accompanying latitude-longitude coordinate pair. These can typically be provided by GPS-enabled devices or by dragging a photo onto a map interface[2]. Based on the new image, SpiritTagger assembles visually relevant photos weighted by geographic distance from the input image. A candidate set of tags is then collected, and relevance scores for *geographically representative* tags seen frequently in similar images are boosted. The highest scoring tags can then be suggested to the user for annotating her upload.

The rest of the paper is organized as follows. Section 2 describes general tag suggestion using mining techniques. Section 3 explains SpiritTagger's filtering and reranking for geographically aware tag importance. A discussion of experiments comparing SpiritTagger's performance to an image similarity baseline is provided in Section 4. Finally, Section 5 provides a summary and directions for future work.

2. TAG SUGGESTION BY MINING

Many algorithms exist for annotation by applying a computer vision model to images. Work by Malik *et al.*[18] and L. Fei-Fei[15], for instance, fits into this category. However, these algorithms are most successful for clean and object-oriented image sets. Photo-sharing websites, populated by real-world tourist photos, will consist primarily of images with cluttered, natural scenes which pose significant problems for such modelling. In addition, computer vision solutions that use one-keyword, one-model algorithms require significant computation since they require separate model creation and classification decisions for each annotation.

These algorithms also do not utilize the large number of community provided annotations which can provide a way to mitigate some of the aforementioned difficulties. We propose to annotate a target photo by mining the collection for similar photos that offer geographic and visual relevance. The collection itself provides the source for tag suggestions which can then be offered to the user for quick additional annotation through a simple mouse click on the relevant options amongst those offered. While annotation by mining has been suggested in the literature by researchers in both image[16] and video[7], our system also uses additional geographical context information in addition to visual image similarity for annotating geo-tagged photos.

2.1 Geographic Mining

Perhaps the simplest suggestion tool for geo-tagged photos collects common tags from images with associated GPS coordinates within a certain radius of the candidate photo. Geographic mining collects the number of times an annotation is seen for another image within a certain radius r. One pitfall is that without filtering, a single user's annotations over multiple photos can often erroneously share the same GPS coordinate. This commonly occurs when a group of photos are dragged from an album onto a single location in a map interface that facilitates geo-tagging. When this occurs close to a target photo, these noisy tags may dominate the annotation scoring. Therefore, we introduce a limitation, and require that each individual user U_i may "vote" for a particular annotation no more than once. Consider the set of users $\mathbf{U} = \{U_1, U_2, ...U_{|\mathbf{U}|}\}$ who have at least one image in the set of geographically close images for the target photo. Then, the collection of annotations for similar images from each user U_i is used in suggestion, call them $A_i = \{a_1^{(i)}, a_2^{(i)}, ...\}$. The annotation is given a score based on the number of users that used that tag. A tag suggestion tool using geographic radius tag a with the formulation:

$$S(a) = \sum_{i=1}^{|\mathbf{U}|} |a \cap A_i| \qquad (1)$$

Annotations with a high score, meaning many users in geographic proximity have applied those annotations, can be supplied as suggested tags to the user.

2.2 Dual Geographic and Visual Mining

Simple geographic mining does not use the full power of a georeferenced photo set as it ignores the visual features of a photo. Building on the geographic mining formulation, we now integrate image similarity to pare down candidate photos to a visually similar set that are mined for annotations. Such an attack has been proposed and shown to return relevant photos in systems such as MediAssist[11].

In our system, global color, texture, edge features and SIFT[6] local features are extracted for the set of images within a certain geographic radius from the target photo. The N-nearest neighbors are retained, and a scoring system similar to that in the Section 2.1 is performed. We collect the set of unique users $\mathbf{V}$ with the associated user-annotation sets, $\{A_1, A_2, ...A_{|\mathbf{V}|}\}$. An additional similarity term can be included that weights visual neighbors by their geographical distance from the target photo. Using the notation α_v for this visual similarity term which is a function of the visual distance between the target and image i, the annotation score for term a is now:

$$S(a) = \sum_{i=1}^{|\mathbf{V}|} \alpha_v(i) \, |a \cap A_i| \tag{2}$$

Our tool, SpiritTagger, builds on this dual mining method. As described in the next section, tags are boosted when found to be relevant semantically in a wider location such as a city or region.

3. TAG RERANKING AND FILTERING BY SPIRIT

The premise of SpiritTagger is that there exists for geographic regions a set of representative tags which can be derived from their local frequency of use in comparison to their global frequency. We use this premise as a method for reranking tag suggestions in a way that reflects the local spirit of a place and improves relevancy of top returns. Suggestions may be especially useful for tourists as a quick way to annotate their vacation photos with distinctive labels. While previous work utilized tag distributions in a geographic area in order to find representative tags for visualization and knowledge extraction[1], here we go further by using tag distributions to calculate a georelevance weighting scheme for tag suggestion for newly added images in a region.

Tag importance is primarily calculated from the ratio of tag frequencies between the region of interest and globally as measured by the number of unique users of a tag. To illustrate, Figure 3 compares unique tag user frequencies for a set of tags crawled from Flickr in an area restricted to Los Angeles and a set of tags crawled globally. Examples of high frequency tags in LA but with lower frequency globally, such as "cars", "freeway", and "palm", will potentially be granted more importance.

To further ensure the usefulness of the tag ratio information, we add two terms to the importance equation which serve to filter out noise in the tags. The first term, $\varsigma(a)$, penalizes tags that do not occur very often both globally and locally. The second term penalizes tags which correspond with very specific geographic locations by considering the ratio of the standard deviation of geographic coordinates for each use of the tag to the maximum standard deviation for the dataset of any tag a.

The equation for tag importance, α_g, as a function of tag a consists of three linear terms:

$$\begin{aligned}\alpha_g(a) = \log\left(\frac{f_{local}(a)+1}{f_{global}(a)+1}\right) + \lambda_1 \cdot \varsigma(a) \\ + \lambda_2 \cdot \left(\frac{\sigma_{geo}(a)}{\max(\sigma_{geo})}\right)\end{aligned} \tag{3}$$

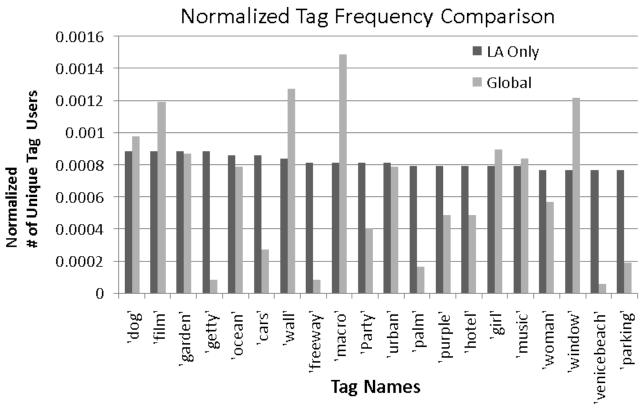


Figure 3: Twenty ordered tags shown to demonstrate tag frequency differences between Los Angeles region and globally. Tags such as "getty", "cars", "freeway", and "palm" with a higher normalized frequency in Los Angeles are weighed more.

with ς, the minimum use penalty term, as:

$$\varsigma(a) = \log\left(\min\left(F_{local}(a), F_{global}(a)\right) + 1\right) \tag{4}$$

f is the frequency as measured by the normalized number of unique users per tag, F is the unique user tag frequency without normalization, λ_1 and λ_2 are weighting factors set at values 0.25 and 0.15 respectively as found to work well for our experiments, and σ_{geo} is the standard deviation of the GPS coordinates of tag a (taken as a sum of latitude and longitude statistics.)

This local tag frequency information is used to further improve the scoring formulation given in the previous section and in equation 1. In particular, the score for annotation a for SpiritTagger now becomes:

$$S(a) = \alpha_g(a) \sum_{i=1}^{|\mathbf{V}|} \alpha_v(i) \, |a \cap A_i| \tag{5}$$

4. EXPERIMENTS

We wish to determine how well the SpiritTagger algorithm for tag suggestion performs compared to baseline methods that use only geographic information or do not rerank tags based on georelevance, and how factors like geographical radius and number of nearest neighbors used vary the results. To do so, we first selected two regions with good coverage: a dense urban section of Los Angeles and the larger region of Southern California. We then crawled a total of 116,281 geo-tagged images from Flickr using their API. 25,988 of these images were randomly selected from anywhere globally, 31,361 were limited to the Southern California geographic region (between 32.5° and 35° latitude and −120.6° and −114.6° longitude), and 58,932 were selected from within the Los Angeles geographic area (between 33.7° and 34.3° latitude and −118.5° and −117.9° longitude). The set of images contained over 48,000 unique tags.

As a test set we selected 99 images from the Los Angeles city data set and 100 images from the Southern California region as candidates for tag suggestion. The images were randomly selected while rejecting images that were overexposed, blurred, or possibly containing privacy concerns.

4.1 Relevance/Coverage

A standard precision-recall metric does not accurately reflect the performance of this algorithm since annotations do not fit neatly into a true/false categorization. For instance, we wish to properly score tags like "Wednesday" which may be correct but not relevant for suggestion. Therefore, we score according to a more appropriate range for annotation ground truth, labeling suggestions as "relevant", "irrelevant," as well as "incorrect." The evaluation metric then provides a tag a with a score, c_a, of $+1$ for a relevant tag, 0 for an irrelevant tag, and -1 for an incorrect tag. A similar three-class scoring method has been adopted for image annotation[16] and video annotation[7].

A modified precision metric, called *relevance*, is defined as the average score in the set $\mathcal{S}$ of extracted tags, $P = \frac{1}{\mathcal{S}} \sum_{a=1}^{\mathcal{S}} c_a$. Additionally, we don't have ground truth for the entire set of correct tags which may be unlimited for photos taken in the world, and therefore a standard recall metric cannot be used. Instead, a running list is kept of all "relevant" annotations for the images encountered through any of the experimental methods and also from the user's own tags. Then, we adopt a recall-like metric, called *coverage*, that indicates the percentage of all seen positive annotations $\mathcal{A}$ covered by the method: $R = \frac{|\mathcal{S} \cap \mathcal{A}|}{|\mathcal{A}|}$, where $\mathcal{S}$ is the set of tags suggested using the particular method. The best metric has the greatest area under the relevance/coverage curve, exhibiting high relevance without expending coverage.

4.2 Groundtruth Annotation

The groundtruth annotation was done by a team of 15 annotators, nearly all of whom were totally unfamiliar with SpiritTagger. To collect ground truth, a web-based tool presented a random photo from the test set along with a Google street map centered at the test image's GPS coordinate. Annotators were then asked to score 10 tag suggestions provided by SpiritTagger, the baseline methods, or the user-supplied annotations. The 15 annotators were instructed to label each suggested tag as either "relevant," "irrelevant," "incorrect," or "unsure." Instructions to the annotators included guidelines such as: a) place names can be relevant if correct b) phone numbers and people's names are irrelevant. Annotators could reference the web in order to determine if an annotation was correct. Tags labeled as "unsure" were not used in scoring the experiments. In total, 16,540 groundtruth annotations were collected which scored approximately 89% of the annotations found by SpiritTagger using any parameter setting.

In order to judge the algorithm's ability to suggest tags that are geographically relevant, we compared the tag suggestions to those generated by a geographic baseline and a dual geographic/visual baseline.

4.3 Geographic Baseline

The geographic baseline collects photos within a certain geographic radius, as described in Section 2.1. Experiments were performed over a radius of 10km, 1km, 100m, and 10m. Results in Figure 4 show a tradeoff between relevance and coverage when varying the radius. A large radius provides greater coverage, since many keywords can be discovered when using a greater geographic range. On the other hand, a small radius such as 10 or 100 meters shows larger relevance since the included photos are likely to contain applicable tags. However, a smaller radius will not cover the breadth of keywords that a large radius over a more diverse set of images will. Another observation is that there is not an increase in possible coverage when using a 10km radius over a 1km radius, shown in Figure 4 as the same coverage endpoint of 0.95, suggesting that correct tags are not found outside of a 1km radius.

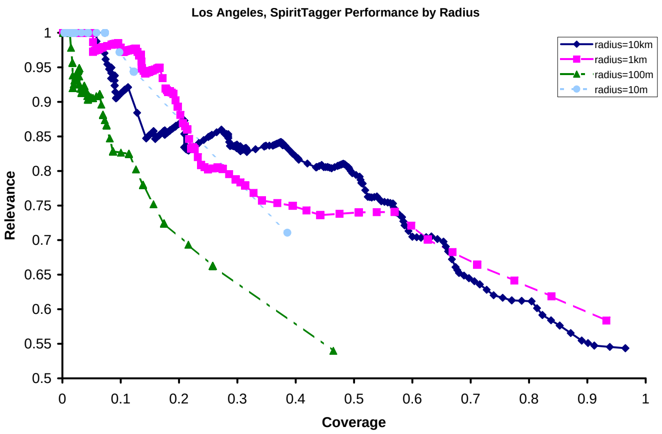


Figure 4: Graph showing performance of geographic baseline by size of radius. Expectedly, using a small radius (10m) has high relevance but does not show the coverage larger radii would. A large radius (1km or 10km) provides better coverage.

We also tested performance using a formulation that rather than taking all tags within a radius, scored them with a exponentially decreasing weighting element for candidate tags based on increasing geographic distance. However, we found this caused a slight decrease in performance. We believe this may be due to noise in the degree of exactness with which people assign GPS coordinates to their photos.

4.4 Visual Baseline

The visual baseline finds close visual neighbors within a certain geographic distance, as describe in Section 2.2. Similarity was formulated using a metric exponentially decreasing with increasing image feature space distance, namely,

$$s_{ij}^{y} = \sum_{y} \exp(-\frac{d_y(x_i, x_j)}{\sigma_y}). \qquad (6)$$

using a late fusion of feature vectors, y, that have been normalized to unit value standard deviation along each dimension. For the similarity measure, we set the decay constant σ_y to the standard deviation of the distance metric used for that feature as seen in the experimental data. Overall similarity between images i and j is given by a weighted linear combination of feature similarities as $s_{ij} = \sum_y \alpha_y s_{ij}^y$.

The SIFT signature, which showed the best performance for cluttered, natural scenes, was weighted the highest. This visual baseline was tested using various similar image set sizes, $N = 20, 10, 5$ for Equation 2. Evaluations show that performance improves when a large number of images, $N = 20$, is kept as seen in Figure 5.

We use three global features and one local SIFT feature,

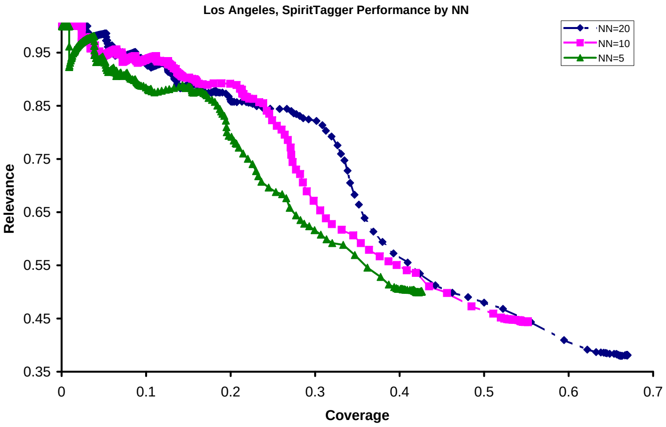


Figure 5: Graph showing performance by number of images, N, contributing tags for suggestion; N in Equation 2. Better performance for higher number of images kept, though some performance loss at points of high precision/low coverage.

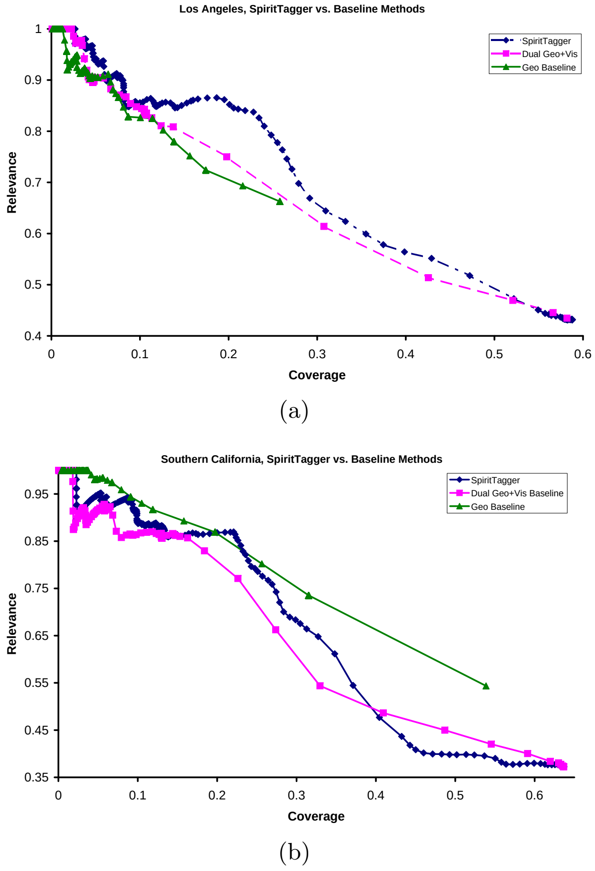
with weights α_f set to 0.033, 0.033, 0.033, and 0.9 by order of enumeration:

1. *Edge Distribution Histogram* 80-dimensional, based on[13].
2. *Homogeneous Texture Descriptor* 48-dimensional, based on[13].
3. *Color Layout Descriptor* 18-dimensional, based on[4].
4. *SIFT Signature* 11,111-dimensional, based on vocabulary tree established by Nister and Stewenius[9] using Lowe's SIFT signature[6] extracted from 5000 random keypoints[10].

4.5 SpiritTagger

The results for the SpiritTagger tool indicate the algorithm's usefulness in an urban area but also show it may be less useful for larger regions. Some examples of the tags supplied by SpiritTagger, using the scoring formulation shown in Equation 5, are shown in Figure 7. The examples show that many of the tags SpiritTagger could have provided were ones actually used by the owner. Additionally, many keywords not used but deemed by evaluators relevant are suggested. Thus, the geo-aware tag suggestion tool aids the annotation process both by making annotation faster (clicking rather than typing) and also by improving coverage of the keywords attached to a photo.

A plot of the relevance/coverage curves for the baseline methods and SpiritTagger are shown in Figures 6(a) and 6(b). One observation is that for the Southern California database, while the performance was better than the dual visual/geographic baseline, it was slightly worse than a simple geographic baseline. The performance loss may imply that the geographic relevance reranking performed by SpiritTagger does not work as well in larger areas. Perhaps this is due to the term that suppresses tags not widely distributed geographically in the region or perhaps it is simply impossible to accurately learn a georelevance score for an annotation that applies uniformly over an entire large region. Additionally, the high performance of the geographic baseline in the Southern California study may be due to high learning of nonspecific but correct tags, such as "LA" or "California" or "USA."



(a)

(b)

Figure 6: (a) Performance of SpiritTagger against the geographic baseline and the dual geographic/visual baseline for Los Angeles dataset. Relevance/coverage shows significant performance improvement over the baseline methods, indicating SpiritTagger's successful evaluation of an annotation's georelevance in order to rerank annotations. (b) Performance of SpiritTagger against the geographic baseline and the dual geographic/visual baseline for Southern California dataset. Relevance/coverage shows slight performance improvement over dual visual/geographic baseline, but does not beat geographic baseline, indicating a limitation in learning annotation georelevance for a large region.

5. CONCLUSIONS

We have presented an algorithm that properly weights geographically relevant annotations for tag suggestion for an image database. The best results use a localized urban area (e.g., Los Angeles) to determine the relevance of a particular annotation to a region, keeps a large number of similar images, and weights the similar images by visual similarity, while discarding precise geographic distance. Selection of geographic radius allows the system to offer a tradeoff between relevance and coverage.

image			
Owner tags (after upload)	trevor gordon, **shortboarding sandspit**	**USA**, Vacation, travel, **California**	**Anaheim**, **Angels**, **minnesota**, **Twins**, **MLB**, **majorleaguebaseball**, **major**, **league**, **baseball**, **ballpark**, *Edison*, **field**, **California**, geotagged
SpiritTagger suggestions (before owner annotation), in order of confidence	**Barbara**, **Santa**, **santa barbara**, **surf**, *Pier*, **surfboard**, *Lemon*, **surfing**, **harbor**, **Surfer**, Wednesday, **channel islands**, **al merrick**, **boarding**, *maassen*, *bogus*, *kneeboard*, **offshore**, *white*, *cockail*, *longboarding sandspit*, *hama*, *tyler anderson*, *damncool*, *oysters*, **person**, *Media*, *Hurricane*, picture, dean, **wave**, **fun**...	**California**, **USA**, **Beverly Hills**, **Stores**, *Archipel*, **90210**, *flagship stores*, **N. Rodeo Drive**, **boutiques**, **designer boutique**, **high end retail stores**, 433 N Rodeo Drive, La Perla, La Perla boutique, *www.laperla.it*, *Rodeo Collection*, Hermes of Paris, 2007, Vacation, **Shopping**, *architecture*, travel	**Anaheim**, **California**, **Angels**, **baseball**, **Twins**, **AngelStadium**, **MLB**, **major**, *Motorcycle*, *jones*, *Edison*, **majorleaguebaseball**, kennedy, *figgins*, *supercross*, *pierzynski*, *ama*, **league**, davanon, **ballpark**, *Racing*, geotagged, **minnesota**, **field**

Figure 7: Example Flickr uploads, the tags the owner ultimately applied for the image, and the tags that could have been suggested by SpiritTagger (bold for correct, plaintext for irrelevant, italics for incorrect.) Note that SpiritTagger has no knowledge of the owner's tags but suggests many of the annotations the owner eventually gives. The left example shows the power of SpiritTagger to properly weight terms particularly applicable to southern California, such as "surf" and "surfboard." Middle example shows SpiritTagger's learning of upscale shopping in Los Angeles. Surfing and shopping are two associations that go beyond place or neighborhood labeling.

Future work will include studies on finer geographic levels to determine how well the algorithm performs for varying levels of scale. A study on annotation specificity, perhaps by adjusting the evaluation relevance/coverage score by inverse document frequency that is inverse annotation use in the database, may better reflect the usefulness of a tag suggestion tool.

6. ACKNOWLEDGMENTS

This work was supported by NSF IGERT Grant #DGE-0221713. The authors would also like to thank Jiejun Xu for his help with experimental logistics.

7. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. H. Yang. World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections. In *Proceedings of the 2007 Conference on Digital Libraries*, pages 1–10, New York, NY, USA, 2007. ACM Press.

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

[3] J. Hays and A. A. Efros. IM2GPS: Estimating Geographic Information From a Single Image. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[4] E. Kasutani and A. Yamada. The MPEG-7 Color Layout Descriptor: A Compact Image Feature Description for High-Speed Image/Video Segment Retrieval. In *IEEE International Conference on Image Processing*, pages 674–677, 2001.

[5] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How Flickr Helps Us Make Sense of the World: Context and Content in Community-Contributed Media Collections. In *Proceedings of ACM Multimedia*, pages 631–640, New York, NY, USA, 2007. ACM.

[6] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.

[7] E. Moxley, T. Mei, X.-S. Hua, W.-Y. Ma, and B. Manjunath. Automatic Video Annotation Through Search and Mining. In *International Conference on Multimedia and Expo*, June 2008.

[8] M. Naaman, A. Paepcke, and H. Garcia-Molina. From Where to What: Metadata Sharing for Digital Photographs with Geographic Coordinates. In R. Meersman, Z. Tari, and D. C. Schmidt, editors, *CoopIS/DOA/ODBASE*, volume 2888 of *Lecture Notes in Computer Science*, pages 196–217. Springer, 2003.

[9] D. Nister and H. Stewenius. Scalable Recognition with a Vocabulary Tree. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2168, Washington, DC, USA, 2006. IEEE Computer Society.

[10] E. Nowak, F. Jurie, and B. Triggs. Sampling Strategies for Bag-of-Features Image Classification. In *European Conference on Computer Vision*. Springer, 2006.

[11] N. O'Hare, C. Gurrin, G. Jones, and A. F. Smeaton. Combination of Content Analysis and Context Features for Digital Photograph Retrieval. In *2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, pages 323–328, 2005.

[12] Panoramio. http://www.panoramio.com/. Website.

[13] P. Salembier and T. Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface.* John Wiley & Sons, Inc., New York, NY, USA, 2002.

[14] C. Shirky. Ontology is Overrated – Categories, Links, and Tags.

[15] G. Wang, Y. Zhang, and L. Fei-Fei. Using Dependent Regions for Object Categorization in a Generative Framework. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1597–1604, Washington, DC, USA, 2006. IEEE Computer Society.

[16] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma. AnnoSearch: Image Auto-Annotation by Search. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1483–1490, 2006.

[17] Webshots. http://www.webshots.com/. Website.

[18] H. Zhang, A. C. Berg, M. Maire, and J. Malik. SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2126–2136, Washington, DC, USA, 2006. IEEE Computer Society.