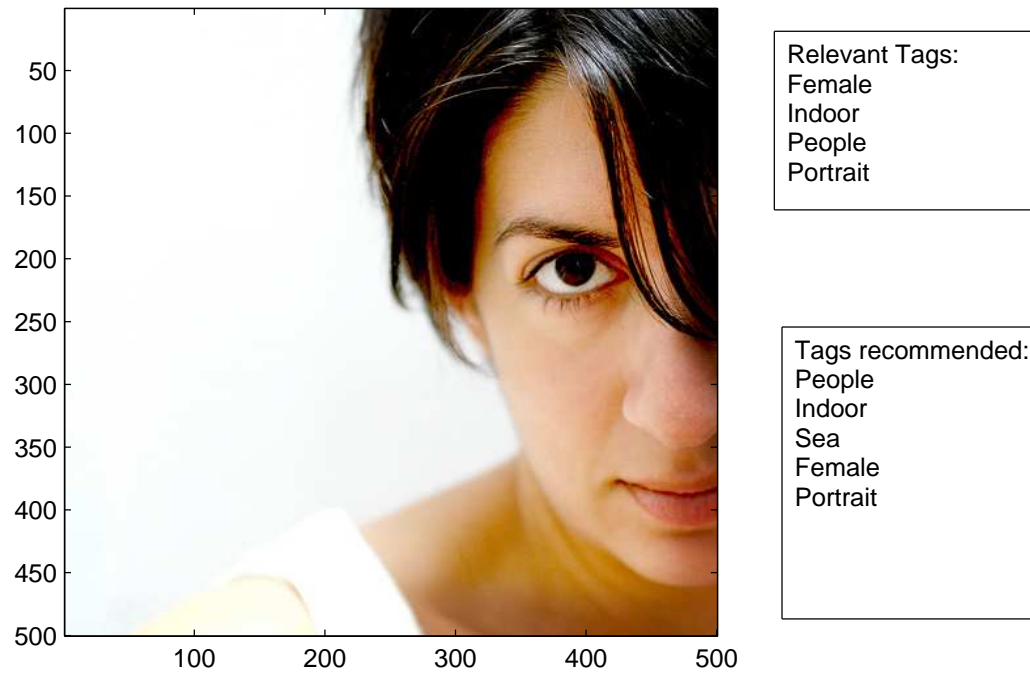


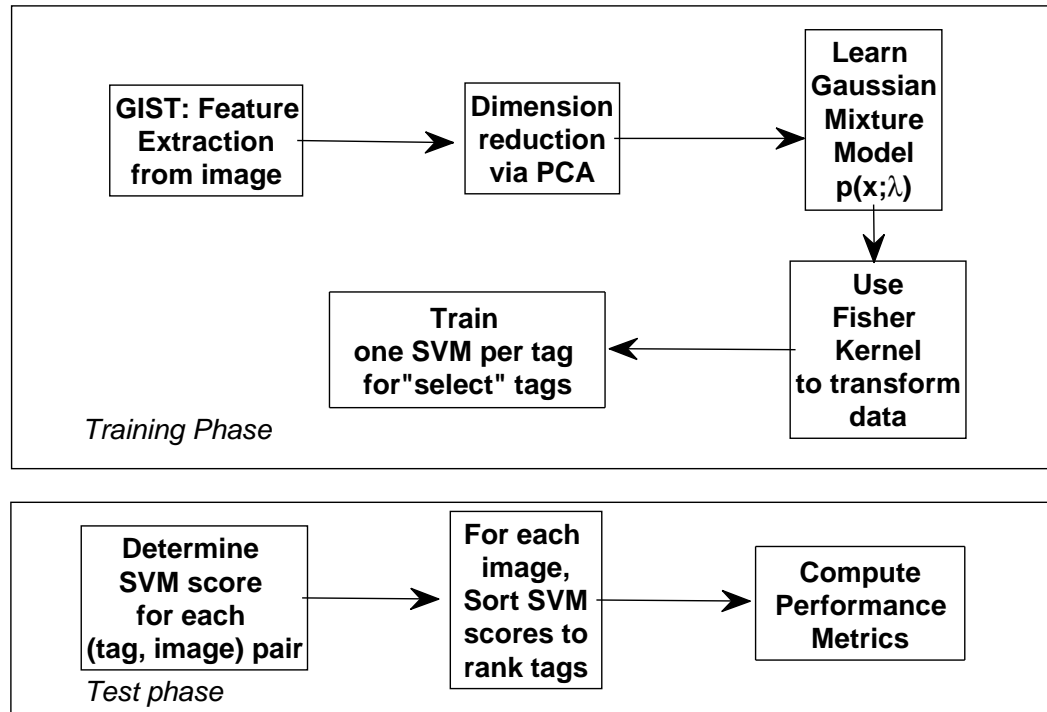
Tag recommendation for photos

What is tag recommendation?

Example: Tag Recommendation system in action



Our algorithm: Block diagram



Tag selection via mutual information [1]

- Only those tags that have a high mutual information with the extracted visual features should be considered by the algorithm.
- To compute joint probabilities for mutual information:
 1. Cluster the image features via K-means.
 2. For each tag j , determine the empirical joint probabilities of the clusters and the presence/absence of tag j .
- Select tag j if the corresponding mutual information exceeds a threshold.

Feature Extraction and Preprocessing

- We start with a dataset of 25000 images and 24 potential tags [2].
- To extract features, we make use of GIST [3], an algorithm that models the dominant spatial structure of a scene.
- Then, we reduce the dimensionality of the data (to 50 features per image) using PCA.
- Next, for each tag j and each y_j , we fit to the training data, a GMM $p_j(x|y_j; \lambda)$ with k components, where x is the PCA-preprocessed image features and $y_j \in \{0, 1\}$ indicates the presence or absence of tag j .

Fisher kernel via diagonal approximation to fisher information matrix

- The fisher kernel function is given by

$$g(x) = F^{-1/2} \nabla_{\lambda} \log p(x; \lambda)$$

where $F = E_X (\nabla_{\lambda} \log p(X; \lambda)) (\nabla_{\lambda} \log p(X; \lambda))^T$ is the fisher information matrix.

- We approximate F by retaining only the diagonal elements as suggested in [4].
- Discriminative training (SVM) is done on the gradient. Note: We implement 1 GMM and 1 SVM per tag.

Performance Metrics

- The algorithm sorts the scores for each tag and outputs a ranked list of tags for each test image.
- **Mean Reciprocal Rank (MRR)**: Average of the reciprocal rank of the first relevant tag in the ranked list.
- **Success@K**: Probability that atleast one among the K tags ranked K or better is relevant.
- **Precision@K**: Probability that a tag ranked K or better is a potential tag.
- We report MRR, Success@1, Success@5 and Precision@5 for our vision-based algorithm and compare it with the reference algorithm presented in [5].

Results

Our vision-based algorithm compares fairly well with the language-based algorithm in [5].

—	MRR	Success@1	Success@5	Precision@5
Our Algorithm	0.67	0.53	0.86	0.38
Reference Algorithm	0.79	0.70	0.95	0.52

References

- [1] E. Moxley, J. Kleban, J. Xu, and B. S. Manjunath, “Not All Tags Are Created Equal: Learning Flickr Tag Semantics for Global Annotation.”, IEEE ICME, New York, Jun. 2009.

- [2] Mark J. Huiskes and Michael S. Lew, “The MIR Flickr Retrieval Evaluation”, In MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval (2008), pp. 39-43.
- [3] Aude Oliva and Antonio Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, International Journal of Computer Vision 42(3), 145175, 2001.
- [4] Perronnin, F., Dance, C., “Fisher kernels on visual vocabularies for image categorization.”, Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- [5] Brkur Sigurbjrnsson and Roelof van Zwol, “Flickr Tag Recommendation based on Collective Knowledge”, WWW 2008 / Refereed Track: Rich Media.