A Modified Saliency-based Gist Descriptor for Scene Classification

Yao Lu, Liki Liu
June 2016

1. Methods

- **Feature Map**
  We use the same architecture with Saliency Model proposed by Itti et al.[7]
  We linearly filter the image into three sub-channels: color channel, intensity channel and orientation channel.
  Using spatial pyramids with Gaussian filter to generate images in different scales.
  Using the interpolation to calculate the center-surround feature map.
- **Gist Calculation**
  For each calculated image in each sub-channel, made the 4x4 grid.
  Siagian&Itti[1]: Calculate the mean of each patch to generate the gist descriptor of the scene
  Our Method: Using the variance instead of the mean to generate the gist descriptor
- **Classification**
  Siagian&Itti[1]: PCA+ICA+Neural Network
  Our Method: PCA+Linearsvm+L1 penalty

2. Result

USC Campus Dataset[1]

USC Campus Dataset	Methods		
	Siagian&Itti[1]	Variance gist	Siagian+variance
ACB	0.86	0.83	0.91
AnF	0.90	0.87	0.94
FdF	0.92	0.89	0.94

21-Landuse Dataset[2]

21-Landuse Dataset	Methods					
	State-of-Art[2]	Siagian&Itti[1]	SIFT[5]+LDA[6]	GIST[3]	Saliency& Sparse Encoding[4]	Our Method
Accuracy	0.79	0.75	0.80	0.72	0.82	0.83

3. Discussion

- **Why we add variance to the descriptor**
  During my experiments and analysis of the classification result, I think the saliency based mean gist extraction method ignores the chaos/entropy of the local patch, which can also be useful for the scene description.
  Besides the variance, I also tested the entropy but didn't work very well.
  I think we can find a more proper way to describe the local patch to achieve higher accuracy.

- **Indoor Scene Classification**
  The gist descriptor drops sharply in the indoor scene in comparisons to other methods. I use the INDECE Indoor dataset for test.
  The state-of-art method[8] achieves accuracy higher than 90% but our methods (mean gist 77%, variance gist 73%, combine 81%) didn't perform well.

4. Future work

- Find a better framework to represent the characteristic of the local patch generated by saliency map

- The saliency based gist descriptor did perform extremely well on satellite images/aerial image, maybe we can do more work to explore why its efficient in the aerial images description and try to achieve better result.

5. Reference

[1] Siagian, C., & Itti, L. (2007). Rapid biologically-inspired scene classification using features shared with visual attention. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(2), 300-312.

[2] Yang, Y., & Newsam, S. (2010, November). Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 270-279). ACM.

[3] Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer *vision*, *42*(3), 145-175.

[4] Zhang, F., Du, B., & Zhang, L. (2015). Saliency-guided unsupervised feature learning for scene classification. *Geoscience and Remote Sensing, IEEE Transactions on*, *53*(4), 2175-2184.

[5] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints.*International journal of computer vision*, *60*(2), 91-110.

[6] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.

[7] Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11), 1254-1259.

[8] Pronobis, A., Caputo, B., Jensfelt, P., & Christensen, H. I. (2006, October). A discriminative approach to robust visual place recognition. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 3829-3836). IEEE.