All 48 Features:

'rrgb', 'rhsv','rlab', 'rrgb_roi', 'rhsv_roi',...
'rlab_roi','rrgb_bg','rhsv_bg','rlab_bg','rgabor',...
'rgabor_roi','rgabor_bg','rhaar','haar_roi','haar_bg',...
'rgist', 'rRgbV3H1_hvecs32','rRgb_hvecs32','rLabV3H1_hvecs32','rLab_hvecs32',...
'rHsvV3H1_hvecs32','rHsv_hvecs32','rHarrisSiftV3H1_hvecs','rHarrisSift_hvecs','rHarrisHueV3H1_hvecs',...
'rHarrisHue_hvecs','rGist_fvec','rDenseSiftV3H1_hvecs','rDenseSift_hvecs','rDenseHueV3H1_hvecs',...
'rDenseHue_hvecs','ropponentsift4096','rcsift4096','rhuesift4096','rsift4096',...
'rrgbsift4096','rrgsift4096', 'rsift512', 'rhuesift512','ropponentsift512', ...
'rrgsift512','rcsift512','rrgbsift512','rhsvsift512','rsift1000',...
'rhuesift1000','rcsift1000','ropponentsift1000'};

Holistic Image Features Extraction for Automatic Image Annotation

Mungkol @ KAMEYAMA Lab.,
2010.08.08

Considerations

- With more computing time and adaptive fitness of

Contents

- Background
- Methodology
- Results and Considerations

For effective and efficient search over billion of images, each one of them needs some meaningful description (annotation).

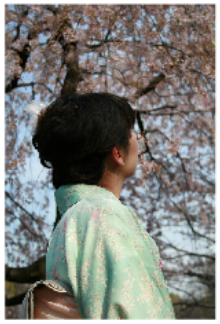
Image Annotation

Given an image, what are the words that describe the image?

For effective and efficient search over billion of images, each one of them needs some meaningful description (annotation).

Image Annotation

Given an image, what are the words that describe the image?



Model

Human Annotation

is

Expensive

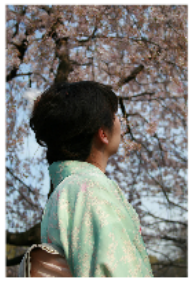
Annotation

Woman, Kimono,
Sakura, Hanami

For effective and efficient search over billion of images, each one of them needs some meaningful description (annotation).

Image Annotation

Given an image, what are the words that describe the image?



Model

Human Annotation

is

Expensive

practical method

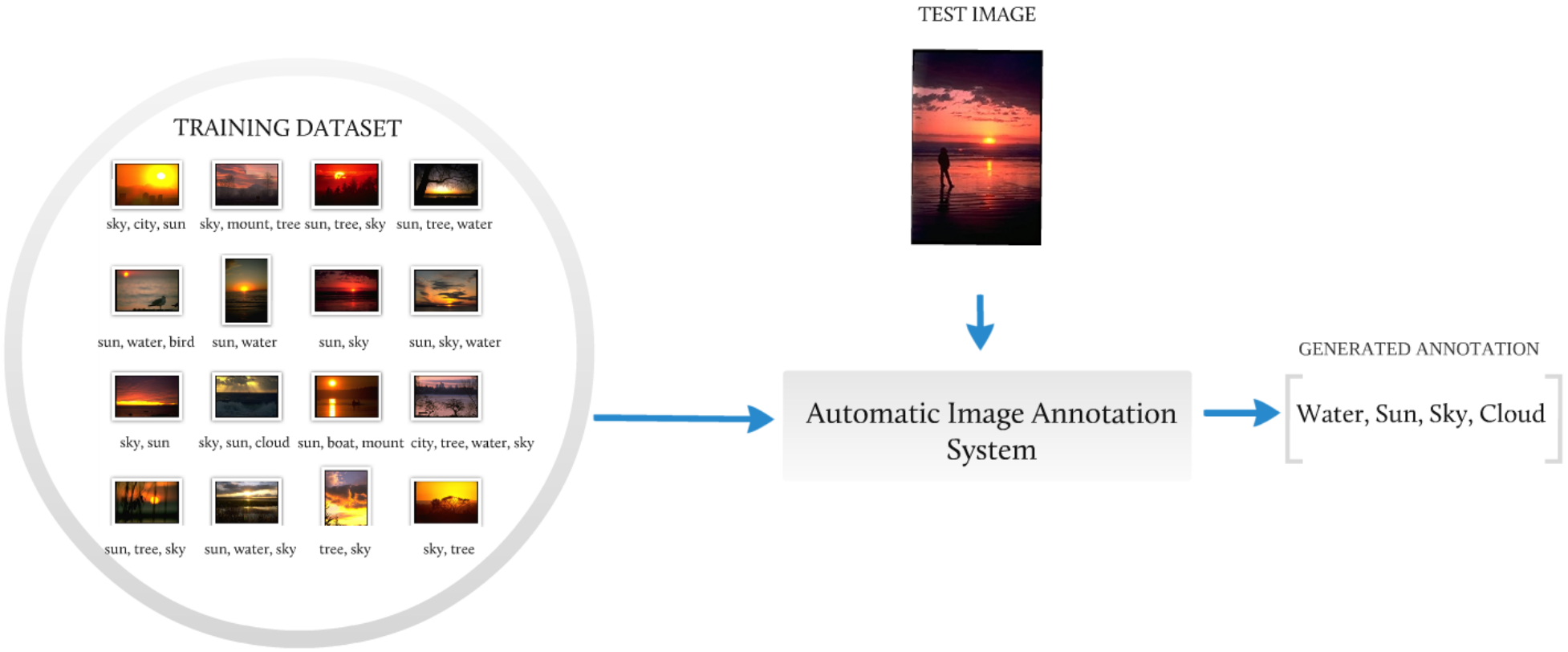
Automatic Annotation

Annotation

Woman, Kimono,
Sakura, Hanami

?????

Automatic Image Annotation System - The concept



Our Proposed Method

Holistic Features Extraction to capture semantic

- Scene level: global + background
- Subject level: saliency regions

Simple annotation scheme using
K Nearest Neighbors (KNN)

Details on Features Extraction

Extracted features:

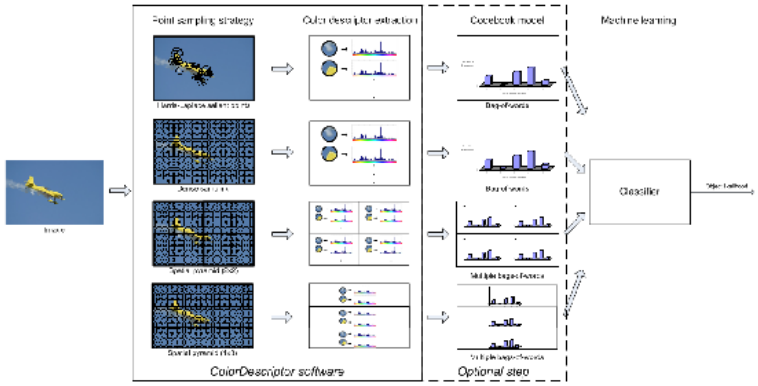
- color: rgb, hsv, lab
- texture: gabor, haar
- scene: gist

[4] Torralba et al. Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision, Vol. 42 (3), 2001

- local saliency
  - color: rgb, hsv, lab
  - texture: gabor, haar

[2] X. Hou and L. Zhang. Saliency detection: a spectral residual approach. In proc. of International Conf. on Computer Vision and Pattern Recognition 2007 (CVPR 2007)

[3] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk, Frequency-tuned Salient Region Detection, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2009.



Created by Koen van de Sande
© University of Amsterdam

Available from
http://www.colordescriptors.com

- Local SIFT features
  - Extract on dense multi-scale grid, and interest points
  - K-mean quantization in 1000 visual words
- Local Hue features
  - Extract on dense multi-scale grid, and interest points
  - K-mean quantization in 100 visual words

* Spatial 3 x 1 partitioning
  - Concatenate histogram from regions

- Local color SIFT features
  - Extract on dense multi-scale grid
  - Features are: rgbsift, opponentsift, huesift, hsvsift, csift, huesift

Annotation Details

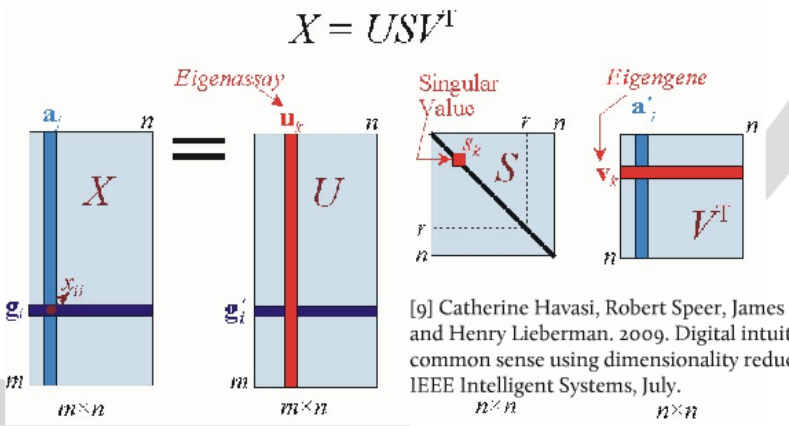
1. Rank Normalization

2. L1 distance

3. KNN by Joint Equal Combination (JEC [1])

[1] A. Makadia, V. Pavlovic, and S. Kumar. A new baseline for image annotation. In proc. of European Conference on Computer Vision, 2008 (ECCV 2008)

4. Relationship between keywords
   - a sparse matrix of global co-occurance frequency
   - singular value decomposition (SVD)



[9] Catherine Havasi, Robert Speer, James Pustejovsky, and Henry Lieberman. 2009. Digital intuition: Applying common sense using dimensionality reduction. IEEE Intelligent Systems, July.

Feature selection

- Since the features are somehow redundant, I use the following method to select a subset of features

Reduce & Enlarge

```
setF = {f1, f2, ..., f31}
sub_setF={}
While setF<>sub_setF
    score= calculate Fmeasure using all features in setF

    For each fi in setF
        score_fi=calculuate Fmeasure using features in (setF -fi)
    End

    to_be_removed_features = fi that has score_fi < score
    setF=setF - to_be_removed_features

    score=calculate Fmeasure using all features in setF

    For each fi in to_be_removed_features
         score_fi=calculate Fmeasure using features in (setF - fi)
    End

    to_be_added_features = fi that has score_fi>score
    sub_setF = setF + to_be_added_features
End
```

Experiment Setting

Corel5k Dataset

- 5000 images (train=4500)
- Vocabulary size = 260
- Words per image = 3.5 (max=5)
- Images per word = 58.4 (max=1004)

- Evaluation/Measure

- Recall
- Precision
- Number of recalled keywords

Experiment Result

- Equal summation of all the selected features

Precision	Recall	Fmeasure	Nb. of KW
0.3116	0.3924	0.3474	161.00

Vs.

0.33	0.42	0.3696	160.00

State-of-the-art performance

[1] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In ICCV, 2009

Considerations

- With more computing time and adaptive fitness of distance, the result can be better
- For instance by fitting the test dataset using Genetic Algorithm to find the coefficient of linear model that maximize Fmeasure

$$d(i,j) = \sum_{k=1}^{N} w_k \tilde{d}^k_{(i,j)}$$

Precision	Recall	Fmeasure	Nb. of KW
0.3116	0.3924	0.3474	161.00
0.3330	0.3979	0.3626	167.00
Vs.			
0.33	0.42	0.3696	160.00

- However, earlier result by fitting the training dataset does not give better performance. This is because we maximize the Fmeasure of the tranning data overally, therefore, the best coeffient of the tranining dataset might not be the best for the test set

Conclusion

- With some variation and new features, we might get similar or hopefully better results than the state-of-the-art but the problem is the computing time which is not scalable.

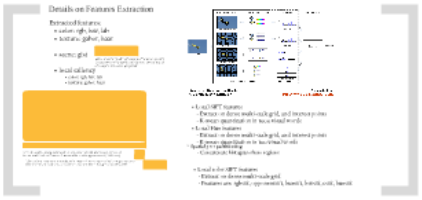
- I am working on extracting more features from the saliency region and background and identifying important features.

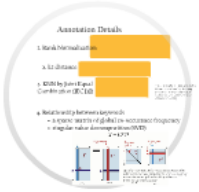
- I also plan to use Evolutionary Programming instead of Genetic Algorithm so that it can better fit individual image rather than the whole dataset. However, the algorithm will still be very expensive.

THANK YOU !

for your kind attention

Q/A?





Contents

- Background
- Methodology
- Results and Considerations

Holistic Image Features Extraction for Automatic Image Annotation

Mungkol @ KAMEYAMA Lab.,
2010.08.08