Group for medical informatics
University Hospital Geneva

Internship 01.07.06 - 31.10.06

Tobias Gass

22.01.2007

Description of the group

- Mainly involved in international health care projects
- Standardizing data
- Building databases with medical annotations
- Only one doctorate candidate actually involved in image processing/classification
- Identification of lung diseases

General Situation:

- Large office in former bank, with air condition (!)
- Most people there worked on encryption of critical data
- One (old) pc, first two weeks without network
- One shared 2.8 ghz machine to work on
- Not too much qualified people to talk to
- The GIFT!

GNU Image Finding Tool

- Framework for Content Based Image Retrieval
- Client/Server architecture using MRML as communication protocol
- Bad documentation!
- Mostly c/c++ implementations with perl as scripting language
- Functionality inspired by information retrieval
- Frequency based scoring on simple, sparse features
- Works well when user feedback is considered

Task:

- Improve classification performance (or look at limitations of GIFT)
- Increase usage of training data

GIFT Features

- Four groups of features
- Color features:
  - Global color histogram (166 bins)
  - Sparse binary color features representing the mode color in predefined blocks(166 possible bins per block, 340 blocks)
- Texture features
  - Global gabor histogram (120 bins)
  - Local gabor histograms of the smallest available blocks (120 possible bins each, 256 blocks)
- 87,446 possible features in total (coloured pictures usually have $O(10^3)$ features)

GIFT scoring

- Score of result image k in respect to query image q:

$$\text{score}_{kq} = \frac{\sum_{j\in\{k,q\}} \text{featureweight}_j^k}{\sum_{i\in q} \text{featureweight}_i^q}$$

- GIFT employs several weighting schemes
- Most common: *term frequency/inverse document frequency* (tf/idf)

$$\text{feature weight}_j^k = \text{TF}_j^k * \log^2(1/(cf_j))$$

- This method only uses the collection frequencies (cf) extracted from training data
- For speed, the mapping from images to features is inverted(Inverted File)

feature $\longrightarrow$ list of images containing that feature

Overview

- Enhancing of color space
- Additional feature weighting
- Weighting of feature groups
- Challenge tf/idf

Enhancing color space

- Most experiments on IRMA db (b/w)
- increased number of grey levels to get more information

Table 1: Error rates on the IRMA database using a varying number of grey levels.

Number of grey levels	Error rate
4	32,0%
8	32,1%
16	34,9%
32	37,8%

- Bad performance probably due to increased sparseness in color block histograms
- Seperately increasing number of grey levels for global and local color features did not help

Additional feature weighting

- Idea: weight features due to their occurance in pairs of images
- Increase weight if feature occurs more frequent in pairs of images of same class compared to pairs of images from different classes
- Can be done on all pairs
- Pruning strategies:
  - Do not use pairs of images which are classified correctly anyways
  - Just use best positive and worst negative pair per image in respect to GIFT score

Table 2: Additional weighting

Used strategy	Frequency weighting
baseline tf/idf	32,0%
all pairs	35,3%
prune too "easy"	33,2%
best/worst pruning	31,7%

Weighting of Feature Groups

- Feature groups in GIFT are weighted equally
- Idea: train weights!
- Compute feature group distances on training data using *leaving one out* evaluation
- Distances as features, class 0 iff images are from different classes, class 1 iff images are in the same class
- Train LDA, use transformation matrix entries as weights
- LDA decreased ER from 32% to 31.7%
- ME!?

Use different scoring methods

Table 3: Error rates on the four feature groups using several weighting approaches.

Feature group	unweighted baseline	with tf/idf	learned weights	tf/idf+learned weights
Color block	36,6%	39,6%	35,1%	40,4%
Color hist	74,5%	–	73,8%	–
Gabor block	56,3%	42,3%	50.0%	45,4%
Gabor hist	53,1%	–	51.8%	–

- Experiments on single feature groups made clear that the weighting mechanisms worked different for feature groups
- $\rightarrow$ use different scorings per feature group
- NOT IN GIFT!
- Wrote small program which implements GIFT features (besides inverted files)
- Considerable decrease in ER 32,0% $\rightarrow$ 27.5% (Euclidean distance: 29.8%)

Setup:

Table 4: Best setup for classification.

Feature group	scoring method	learned feature weights
Color block	L2	–
Color hist	GIFT histogram intersection	–
Gabor Block	GIFT tf/idf	used
Gabor hist	L2	used

Additional results:

- Using the aspect ratio of medical images decreased the ER considerably to 26.4%
- Applying the found method to the non–medical automatic annotation task (ltu database) yielded similar results

Table 5: Error rates on the LTU database using various strategies.

Method used	Error rate
baseline	91,7%
with learned feature weights	90,5%
with mixed scoring	88,3%

Conclusions(technical):

- GIFT is strange!
- Applicability of frequency based weighting schemes seems to depend on features
- Tf/idf(with IF) is fairly fast on large, sparse feature spaces
- GIFT works quite well considering relevance feedback/multiple query images
- Given one query image only, term frequencies are always 1, maybe query expansion might help
- More sophisticated features might help, currently experiments on sparse histograms extracted from image patches are running

Conclusions(other):

- Switzerland is great!
- Lots of nice people
- And mountains!
- Feedback was sparse, too

Thank you for your attention!