Group Based Classification (GBC) for Medical Diagnostics

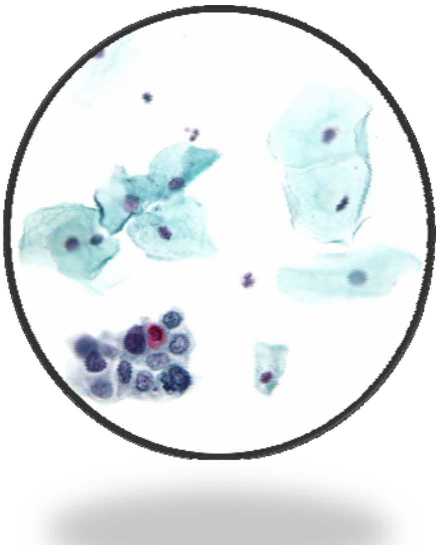
A/Prof. Andrew P. Bradley
Biomedical Engineering
The University of Queensland

Overview

- Where does GBC come from?
  - Cervical Cancer Screening
  - Compound Classification
- What is GBC?
- How to implement GBC?
  - Some *k*-NN based approaches and some results
- Where to apply GBC?
  - Hopefully, a good question

Cervical Cancer Screening

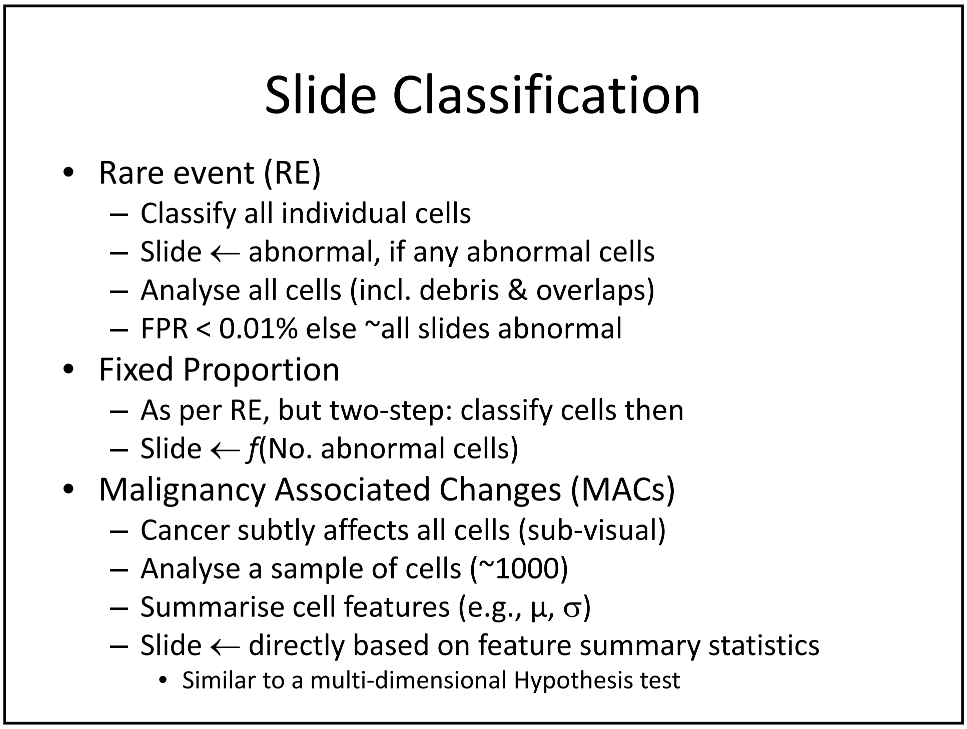
- Based on Pap smear
  - Sample of cervical cells
  - Microscopically analysed
- Aim to detect pre-cancerous changes
- Largest volume cytological test
- One of the “classic” problems in Pattern Recognition



The Challenges

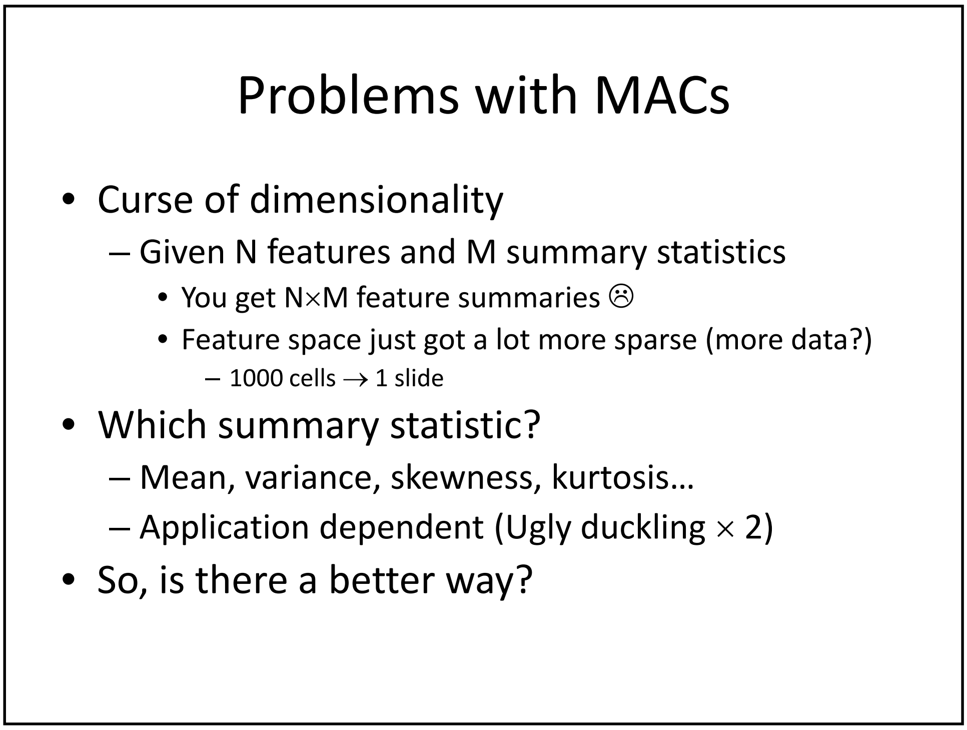
- Image acquisition
  - Automated µscope or slide scanner
- Scene segmentation
  - Detect and segment cell nucleus and cytoplasm
- Features extraction
  - CN ratio, chromatin distribution (nucleus texture), OD etc
- Classification
  - Normal or abnormal

- Giga-pixel image
- ~10,000 cells + debris
- Nucleus small: ∅ ~20 pixels
- Cell or slide?



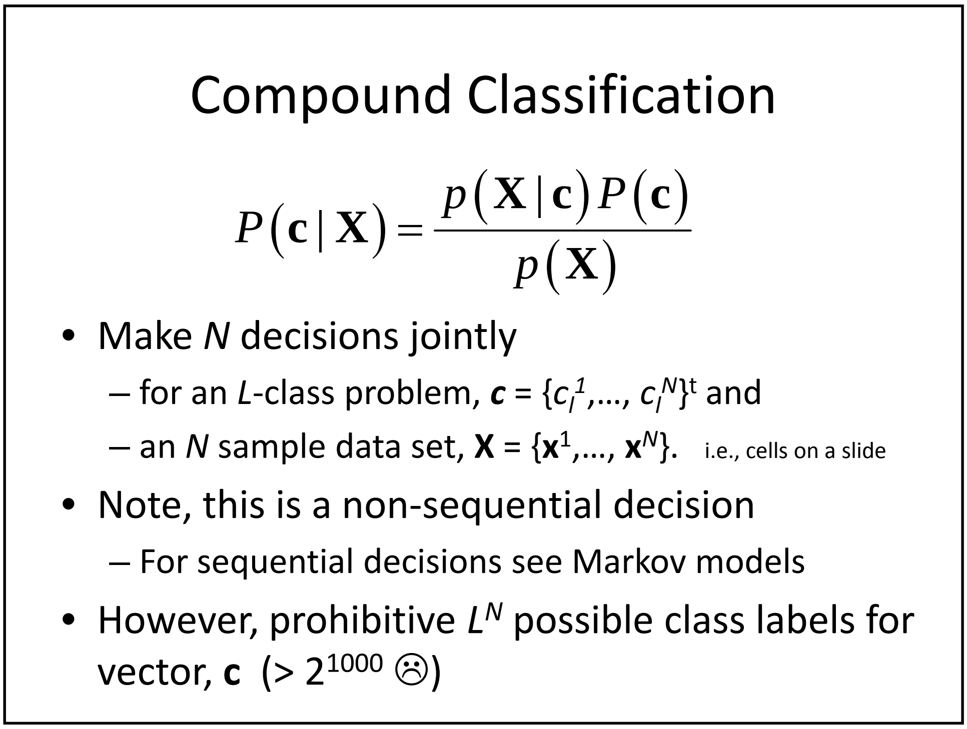
Slide Classification

- Rare event (RE)
  - Classify all individual cells
  - Slide ← abnormal, if any abnormal cells
  - Analyse all cells (incl. debris & overlaps)
  - FPR < 0.01% else ~all slides abnormal
- Fixed Proportion
  - As per RE, but two-step: classify cells then
  - Slide ← *f*(No. abnormal cells)
- Malignancy Associated Changes (MACs)
  - Cancer subtly affects all cells (sub-visual)
  - Analyse a sample of cells (~1000)
  - Summarise cell features (e.g., μ, σ)
  - Slide ← directly based on feature summary statistics
    - Similar to a multi-dimensional Hypothesis test



Problems with MACs

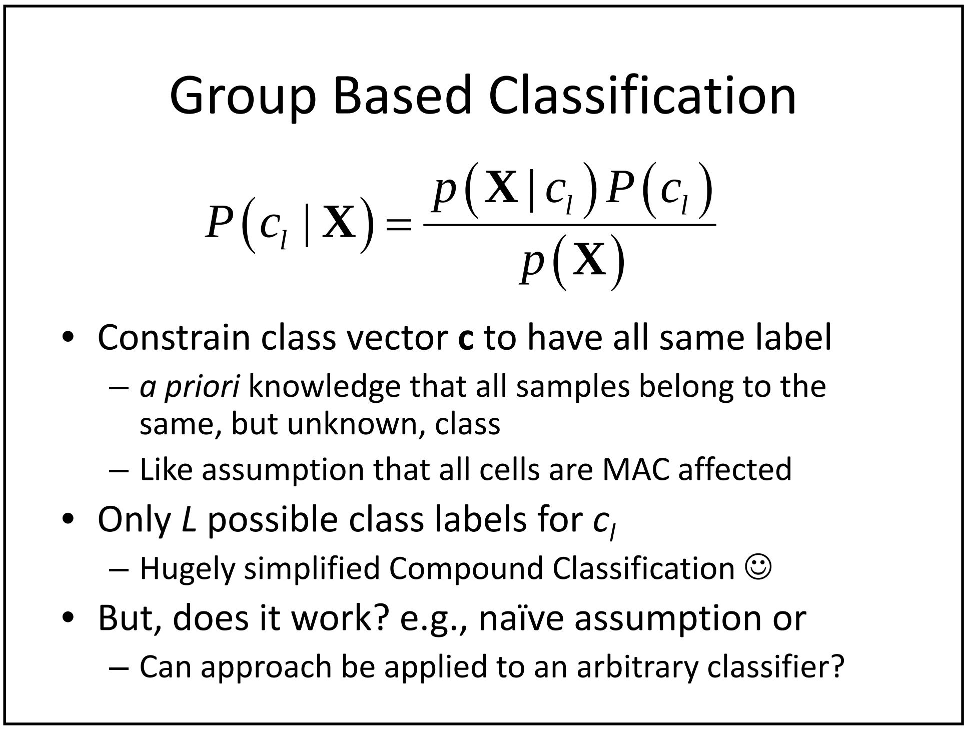
- Curse of dimensionality
  - Given N features and M summary statistics
    - You get N×M feature summaries ☹
    - Feature space just got a lot more sparse (more data?)
      - 1000 cells → 1 slide
- Which summary statistic?
  - Mean, variance, skewness, kurtosis…
  - Application dependent (Ugly duckling × 2)
- So, is there a better way?



Compound Classification

$$P(\mathbf{c} \mid \mathbf{X}) = \frac{p(\mathbf{X} \mid \mathbf{c})P(\mathbf{c})}{p(\mathbf{X})}$$

- Make *N* decisions jointly
  - for an *L*-class problem, ***c*** = {c_l^1,…, c_l^N}^{t} and
  - an *N* sample data set, **X** = {$\mathbf{x}^1$,…, $\mathbf{x}^N$}. i.e., cells on a slide
- Note, this is a non-sequential decision
  - For sequential decisions see Markov models
- However, prohibitive L^N possible class labels for vector, **c** (> 2^{1000} ☹)

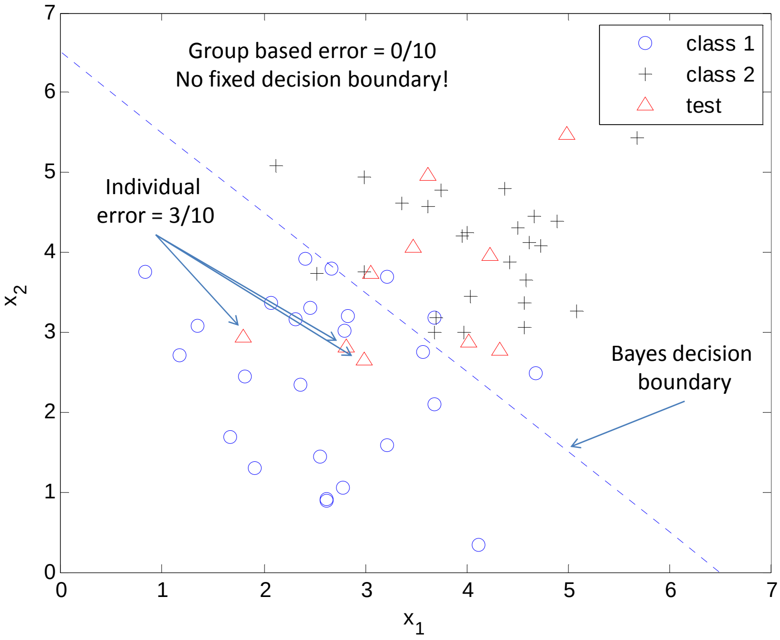


Group Based Classification

$$P(c_l \mid \mathbf{X}) = \frac{p(\mathbf{X} \mid c_l)P(c_l)}{p(\mathbf{X})}$$

- Constrain class vector **c** to have all same label
  - *a priori* knowledge that all samples belong to the same, but unknown, class
  - Like assumption that all cells are MAC affected
- Only *L* possible class labels for c_l
  - Hugely simplified Compound Classification ☺
- But, does it work? e.g., naïve assumption or
  - Can approach be applied to an arbitrary classifier?

An Illustrative Example



Group based error = 0/10
No fixed decision boundary!

Individual error = 3/10

Bayes decision boundary

Group Based *k*-Nearest Neighbours

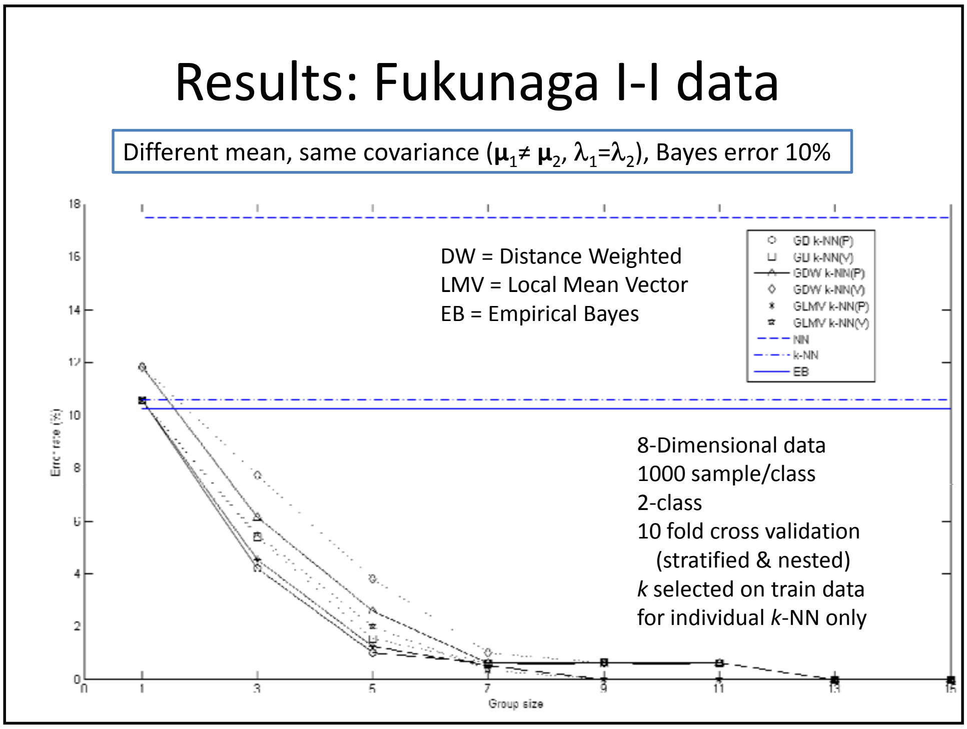
e.g., 3-NN classifier, test set “group” size 5, binary class {-1, 1}

Test set	Class of nearest neighbours			Individual 3-NN
t_1	1	1	1	1
t_2	-1	-1	-1	-1
t_3	1	-1	1	1
t_4	-1	-1	-1	-1
t_5	1	1	-1	1

Row-wise voting
2:3 → 1
GB *k*-NN(V)

Column-wise pooling
8:7 → -1
GB *k*-NN(P)

a priori knowledge: all test set has same class

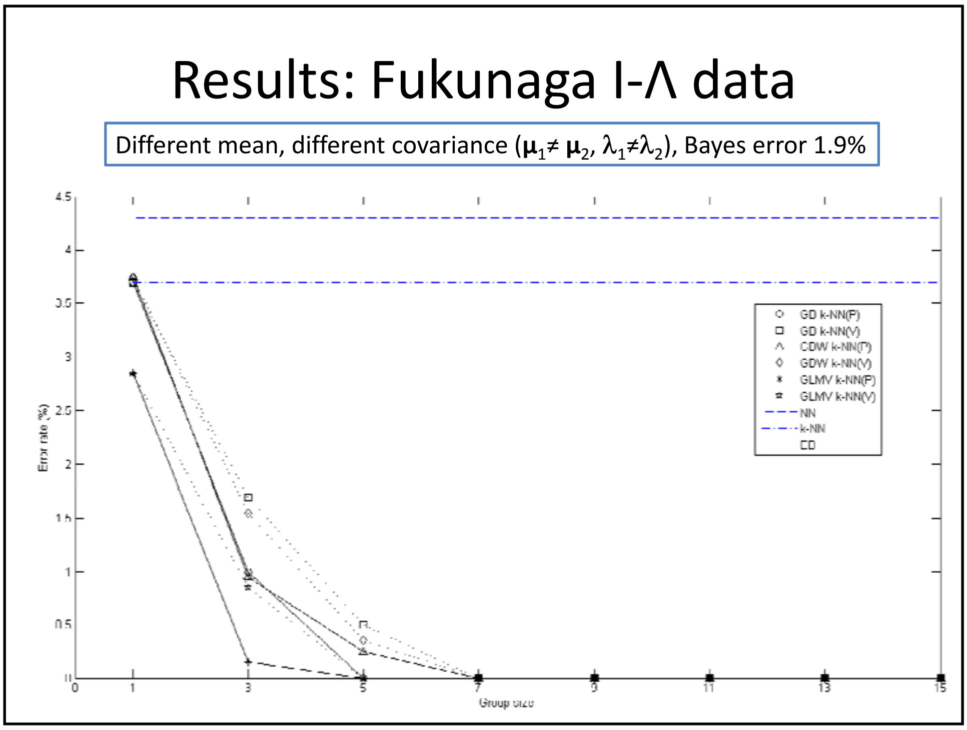


Results: Fukunaga I-I data

Different mean, same covariance ($\mu_1 \neq \mu_2$, $\lambda_1 = \lambda_2$), Bayes error 10%

DW = Distance Weighted
LMV = Local Mean Vector
EB = Empirical Bayes

8-Dimensional data
1000 sample/class
2-class
10 fold cross validation
(stratified & nested)
k selected on train data
for individual *k*-NN only

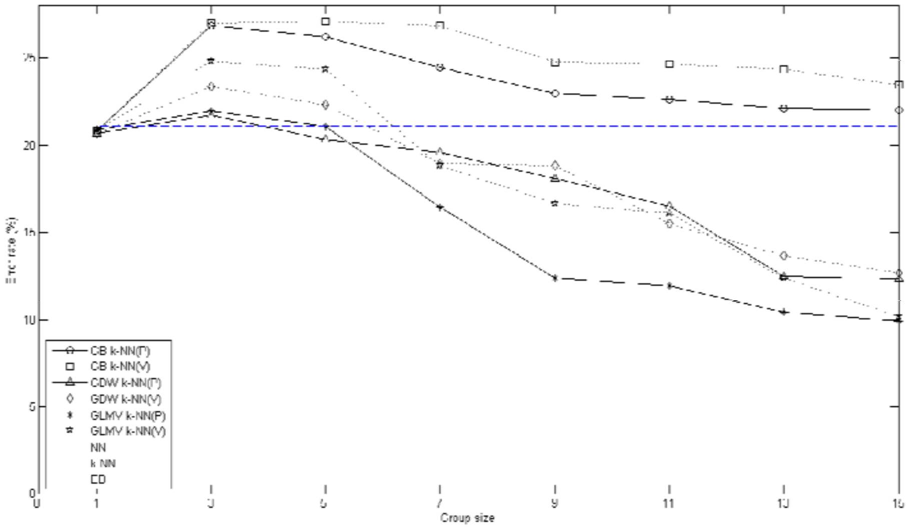


Results: Fukunaga I-Λ data

Different mean, different covariance ($\mu_1 \neq \mu_2$, $\lambda_1 \neq \lambda_2$), Bayes error 1.9%

Results: Fukunaga I-4I data

Same mean, different covariance ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $\lambda_1 \neq \lambda_2$), Bayes error 9%



Synthetic Data Summary

- Pooling error is typically < voting error (33:3)
  - Indicates that GBC in 1-step is better than
    - 2-step individual classifier plus voting
- Where classes have different means (I-I, I-Λ)
  - Groups of 3 error rate < (individual) Bayes error
  - Increasing group size, error rate → 0
- Where classes only differ in variance (I-4I)
  - *k*-NN not really suited to problem, but
    - Improved by DW and LMV variants
  - Trend of reducing error as group size increases
- But what about real-world data?

Results: Pap Smear Data

- 99 Normal, 40 Abnormal slides (≥ CIN1)
  - 1000 cells per slide, 29 feature vector, MACs (μ, σ)
- Stratified 10 fold cross-validation (test on ~14 slides)
  - Best 3 MACs features (Mahalanobis criterion)
  - *k* selected on MACs, on training data for *k*-NN only
  - GBC use raw selected features, test set size 100

Classifier	Accuracy ± STD	AUC ± STD
EB (MACs)	80.950 ±6.293	0.604 ±0.213
***k*-NN (MACs)**	81.264 ±6.132	0.658 ±0.099
GB *k*-NN (V)	80.659 ±8.194	0.654 ±0.213
GLMV (V)	78.462 ±6.613	0.611 ±0.130
GB *k*-NN (P)	81.923 ±8.751	0.764 ±0.136
GLMV (P)	79.945 ±8.692	0.693 ±0.220

Pooling > voting
Pooling ≈ MACs
but biased to MACs

Applications of GBC

- Pathology and cytology
  - Classify slides not cells
- Neurophysiology: evoked responses
  - Classify individual responses not grand average
- Document classification: “Bag-of-words” model
- Others, please…
- Consider the Iris data
  - Sepal length & width, petal length & width → species
  - Group No. leaves/sepals from same plant then → species
    - Group based classification ☺
- So, any application where
  - you can *a priori* organise your data into groups?
    - Where class unknown, but know group has same class label

Summary

- GBC is inspired by Pap smear screening
  - Not new, just a simplified compound classifier
- Investigated a couple of implementations
  - Variants of *k*-NN (also hypothesis testing)
  - Promising results on some data sets
  - Lots of possible implementations to try!
- Perhaps of use in other applications?
  - Where you can also group your data

For more details see:

Noor A. Samsudin and Andrew P. Bradley, “**Nearest Neighbour Group Based Classification**,” *Pattern Recognition*, 43 (10), pp 3458-3467, 2010 (DOI: 10.1016/j.patcog.2010.05.010)

Noor A. Samsudin and Andrew P. Bradley, “**Group-based Meta-classification**,” *19th International Conference on Pattern Recognition* (ICPR), Tampa Bay, Florida, pp 2256-2259 , December 2008 (DOI: 10.1109/ICPR.2008.4761778)

The End

QUESTIONS?