Algorithm and Data Structure Coursework: PCA Features for R-tree Based Similar Image Search

Qiwei Feng
2011011250, IIIS-10
Tsinghua University
gdfqw93@163.com

Pufan He
2011011307, IIIS-10
Tsinghua University
hpfdf@126.com

ABSTRACT
This paper provides a sample of a LaTeX document.

Keywords
R-Tree, Similar Image, PCA, K-Means

1. INTRODUCTION
Github address: `https://github.com/caiwaifung/lastcourse`.

1.1 Image Search
1.2 Data Structures
1.3 Low Level Features
1.4 PCA
1.5 K-Means
2. DATA
3. FEATURE FINDING
TODO: list feature here (PCA, KMeans, Composite).

4. R-TREE
We use the "rtree alternative package" implementation of R-tree. The wrapper `src/a.cpp` calls methods of provided R-tree class. Run `python src/run.py` to compile and run the program.

5. EXPERIMENTS
5.1 Node Access Numbers
Table 1 lists the node access number in different cases.

TODO: We can see that blablabla

5.2 Performance
Table 2 lists the correctness for different feature. There are 613 queries in total, and the database varies from 1000 images to 5000 images.

Table 1: Node Access Numbers

Method and Feature Num	1000	2000	3000	4000	5000
Color Moment HSV 9	46.11	67.69	88.69	98.24	117.0
PCA 4	37.18	53.31	71.37	76.72	81.83
PCA 8	68.42	107.7	145.8	176.1	208.4
PCA 12	77.95	129.9	174.5	217.8	252.6
PCA 16	82.46	135.4	190.2	236.2	280.0
PCA 20	81.55	137.8	196.4	253.7	302.8
PCA 24	83.11	135.7	192.8	248.0	297.4
PCA 30	123.8	207.4	281.4	351.7	416.4
KMeans 4	16.01	19.76	22.24	23.53	24.71
KMeans 8	17.49	21.83	24.44	25.88	26.66
KMeans 12	20.90	25.76	30.96	34.85	37.43
KMeans 16	22.22	28.40	33.89	37.25	38.14
KMeans 20	25.25	35.26	39.25	39.67	44.48
KMeans 24	20.68	27.81	30.20	33.46	34.79
Composite 25	80.01	136.8	202.4	254.9	305.4

Table 2: Correctness of Different Feature

Method and Feature Num	1000	2000	3000	4000	5000
Color Moment HSV 9	153	174	178	190	195
PCA 4	116	130	133	141	151
PCA 8	158	170	172	176	183
PCA 12	181	190	199	205	208
PCA 16	181	198	205	206	217
PCA 20	185	200	207	213	225
PCA 24	180	194	201	212	221
PCA 30	177	198	205	203	217
KMeans 4	126	132	147	148	147
KMeans 8	124	155	154	155	161
KMeans 12	132	154	158	154	158
KMeans 16	133	154	165	167	173
KMeans 20	132	160	157	159	164
KMeans 24	121	161	153	167	178
Composite 25	201	219	230	235	240

Table 4: Some Typical Commands

Command	A Number	Comments
`\alignauthor`	100	Author alignment
`\numberofauthors`	200	Author enumeration
`\table`	300	For tables
`\table*`	400	For wider tables

Table 3: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
\$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

6. CONCLUSION

7. REFERENCES

[1] M. Bowman, S. K. Debray, and L. L. Peterson. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.*, 15(5):795–825, November 1993.

[2] J. Braams. Babel, a multilingual style-option system for use with latex's standard document styles. *TUGboat*, 12(2):291–301, June 1991.

[3] M. Clark. Post congress tristesse. In *TeX90 Conference Proceedings*, pages 84–89. TeX Users Group, March 1991.

[4] M. Herlihy. A methodology for implementing highly concurrent data objects. *ACM Trans. Program. Lang. Syst.*, 15(5):745–770, November 1993.

[5] L. Lamport. *LaTeX User's Guide and Document Reference Manual*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.

[6] S. Salas and E. Hille. *Calculus: One and Several Variable*. John Wiley and Sons, New York, 1978.