LOW-BITRATE BENEFITS OF JPEG COMPRESSION ON SIFT RECOGNITION

Mohamed Elhoseiny[1], *Bing Song*[2], *Jeremi Sudol*[2], *David McKinnon*[2]

[1] Rutgers University, Department of Computer Science
[2] IPPLEX Holdings Corporation, Computer Vision Group
m.elhoseiny@cs.rutgers.edu, {bing, jeremi, david}@ipplex.com

ABSTRACT

Feature detection and image matching are two important tasks in photogrammetry. Their application continues to grow in a various fields, from simple photogrammetry tasks such as feature recognition, to the development of sophisticated models to deal with bandwidth problems in mobile devices. Due to low bit-rate requirement of the current mobile communication, Mobile Visual Search became a very challenging problem. In this direction, this paper presents important conclusions based on a comprehensive evaluation of SIFT matching performance against various parameters (e.g. JPEG quality/compression in model and test images, image resolution, etc). The main conclusion of the performed experiments is that reducing jpeg quality from 100% to 30% slightly impart the matching performance, while it significantly reduces the communication bandwidth requirement by $\approx 70\%$.

Index Terms— SIFT, JPEG, Image Matching, Mobile Visual Search, Low Bitrate

1. INTRODUCTION

Lowe [1] presented SIFT for extracting distinctive invariant features from images, which is invariant to image scale and rotation. Since then, it has been widely used in many image retrieval systems and Mobile Visual search applications [2, 3, 4]. Furthermore, many papers have been studying encoding techniques of the images to have a lower bit rate [5, 6, 7] . An important aspect, not comprehensively studied in these systems, is the effects of JPEG quality that significantly affects the image size, and how that reflects the matching performance for distant, close, and average distant objects. Another important question is How the performance changes under different image resolutions.

One of the important points, addressed in this study, is the contrast between transmitting features to transmitting images. Although compressing and reducing features size has been under intensive research recently [8, 6], these methods are not proven to be applied to more complicated context. For instance, Vijay's CHOG approach [5, 6], does not handle spatial constraints (compressed version of HOG [9]) . So as an aspect of this study, we report the bandwidth benefits of transmitting lower quality image compared to features. The rest of this paper is organized as follows. Section 2 presents the datasets utilized in the evaluation and the parameter settings. Section 3 presents the common setup for all experiments. Section 4 presents the experiments and discussion . Section 5 presents the conclusion.

2. PARAMETER SETTINGS AND DATASETS

This experimental study on the SIFT recognition is based on various parameter settings that include

1. With/ Without Upsampling of factor 2 (Width, Height)
2. Different Model JPEG qualities [5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%].
3. Different Test JPEG qualities [5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%].
4. Different Test image resolutions (640, 480, and 320).
5. With Noise / Without Noise (Noise means merging model images with a database of 10000 random images, which is about 1.5 GB on Disk).
6. Different Datasets (Covers1, Covers2), detailed in the following subsections.

In order to study these parameters, two datasets were created (Covers1, Covers2). We created these datasets to analyze aspects like the effect of images taken from different distances and angles, which is a typical situation that many of the current systems necessitates.

2.1. Covers1 Dataset

Covers1 is a dataset of forty four book covers. For each book the cover image is retrieved from the web as the model images and printed. Nine images were captured for each model with complex background as shown in Figure 1. The 9 images were picked according to the specifications of distance and angles shown in Table 1. Figure 1 shows sample images for Covers1 dataset.

2.2. Covers2 Dataset

Covers2 is a dataset of the same forty four book covers as the model images. Fifteen images were captured for model with

Distance	Angles
Close (object is about 35% of the image)	0,10,30, 45
Average Distant (object is about 20% of the image)	0,10,30
Far (object is about 10% of the image)	0,10

Table 1: Covers1 dataset specifications

Fig. 1: Covers1 dataset Images. From left to right Model Image (2516 features) , Test image (Close-0 deg, 337/7562 features matched), Test Image (Moderate -30 deg, 151/7847 features matched) , Test Image (Far-10 deg, 82/5051 features matched). For the test images, the detected features are colored in yellow while the matched features are colored in blue and the matched object is surrounded by a red polygon

simpler background (i.e. much less artificial features). Then, fifteen images were picked according to the specifications of distance and angles shown in Table 2. This dataset is more challenging in terms of distances and angles however it has simpler background as shown in Figures 1, 2. Figure 2 shows sample images for Covers2 dataset..

Distance	Angles
Close (object is about 35% of the image)	0,10,30, 45,60
Average Distant (object is about 10% of the image)	0,10,30, 45,60
Far (object is about 1-2% of the image)	0,10,30, 45,60

Table 2: Covers2 dataset specifications

Fig. 2: Covers2 dataset Images. From left to right, Model Image (2516 features) , Test image (Close-60 deg, 9/416 features matched), Test Image (Moderate -45 deg, 32/262 features matched) , Test Image (Far- 60 deg, 0/1848 features matched).

3. COMMON EXPERIMENTAL SETUP

The setup, shared in all our experiments, is defined as follows

1. Image resolution for the basic 44 model Image Images: Max dimension (640). The images were scaled such that the aspect ration is preserved.
2. SIFT Feature Quality: 0.85.
3. SIFT Feature Size: 128.
4. Affine Matching is used to recognize images [1] (exactly in section 7.4 in this paper)
5. 121 datasets were generated on for each quality ratio of Model and Test image (11 JPEG quality rations [5%, 10%, 20%, 30%, 40%,50%, 60%, 70%, 80%, 90% and 100%] applied for both Model and testing images giving 121 combinations).
6. In case that MergeF (i.e. a flag) = True, the number of models in database = 44. Other wise, the total umber of Models = 10044 because 10,000 random image models were added in the experiment..

4. EXPERIMENTS AND DISCUSSION

We performed seven big experiments, each of them includes 121 sub-experiments (11×11 for each JPEG quality combination). The specifications of each experiment are illustrated in Table 3. For each of these experiments Recall, Precision, and Accuracy were computed. Due to the high volume of the results, we present here the most important parts. The complete set of results are presented in the supplementary materials. Hence, Figure 3 presents the recall of all the 7 experiments.

	Dataset	UpSample	MergeF	Res
Experiment 1	Covers1	No	False	480
Experiment 2	Covers1	Yes	False	480
Experiment 3	Covers1	No	True	320
Experiment 4	Covers1	No	True	480
Experiment 5	Covers1	No	True	640
Experiment 6	Covers2	No	True	480
Experiment 7	Covers2	No	True	640

Table 3: Specifications of the Experiments. *Res* refers to the maximum dimension of the test image resolution. Test images were scaled with aspect ratio preserved.

Table 4 presents the database size of the SIFT features stored in KD tree data structure. The table indicates higher database size for low JPEG quality databases (due to many artifacts of low quality), then it decreases for moderate jpeg quality. Eventually, the DB size increases again. The following subsections discuss different aspects of these results.

4.1. Upsampling Effect

In Experiment 1 (Upsampling = False), the average recall and precision (over all the 121 cases) are 82.34% and 99.57% respectively, while in Experiment 2 (Upsampling = True),

	MergeF = False			MergeF = True			
			Covers1 dataset			Cover2 dataset	
JPEG quality	**480**	**480 with Upsample**	**320**	**480**	**640**	**480**	**640**
5	324393	1797212	1443228930	1443291870	1443229018	1443229062	1443229150
10	299189	1686080	1441545354	1441560506	1441545442	1441545530	1441545574
20	289374	1734801	1441172238	1441148078	1441172326	1441172414	1441172458
30	291315	1676574	1441033738	1441038730	1441033826	1441033914	1441033958
40	292930	1623093	1440999422	1441017742	1440999510	1440999598	1440999642
50	292929	1582059	1440981378	1441012266	1440981466	1440981554	1440981598
60	294015	1545387	1440995182	1441005018	1440995270	1440995358	1440995402
70	294602	1509997	1440963498	1440969014	1440963586	1440963674	1440963718
80	295260	1484038	1440977262	1440975806	1440977306	1440977438	1440977482
90	295397	1474941	1440976990	1440958530	1440977078	1440977166	1440977210
100	295397	1481337	1440988426	1440981854	1440988426	1440988558	1440988558

Table 4: Trained DataBase Sizes in bytes (KD trees of features) for each of the seven experiments against each of the JPEG quality ratios

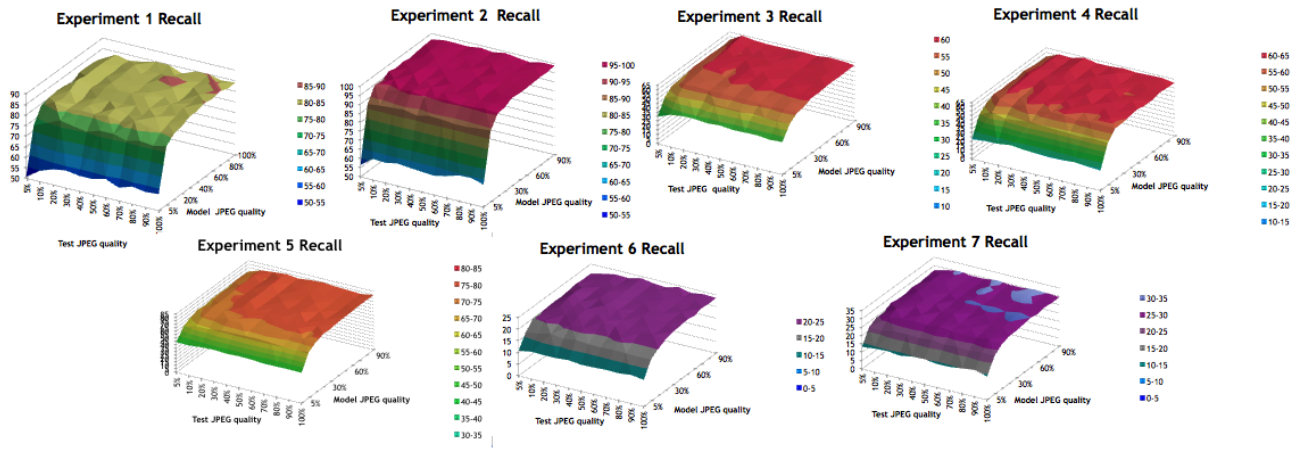

Fig. 3: Experiments' Recall: Each sub-figure visualizes the recall of each of the 121 (11 (model JPEG quality values) $\times$ 11 (test JPEG quality values) sub experiments on a 3D surface

the average recall and precision are 93.25% and 99.21% respectively. From Experiment 1 and 2, the conclusion is that upsampling improves recognition accuracy with about 10% on average with a penalty of high computational time (about three times on average, breaking real time constraints, as detailed in the supplementary materials).

4.2. JPEG Quality Effect

Figure 3 illustrates logarithmic increase behavior in the recognition performance in all of the experiments (1-7), while the JPEG quality ratios change in both dimensions (i.e. model image and test image quality).In other words, There is almost no relative change after certain percentage of JPEG quality ($\approx$ 30%). The bottom-line of these results is discussed here from different perspectives (Image Size effect, Merge Effect, and Test-image Max Dimension Resolution Effect).

4.2.1. Effect on Image Size

Figure 4 shows how the JPEG quality affects the image size on five different cases, which are Covers1 dataset (1-640, 2-480, 3-320 pixels) and Cover2 dataset (4-640, 5-480 pixels). This reflects the bandwidth benefits of using the 30% JPEG quality. Quantitatively speaking, 70% of the bandwidth is reduced in Covers-640 (i.e from 129.5 KB for 90% JPEG quality to 37.9 KB for 30% JPEG quality). This is without losing control on the descriptor size that could be computed on the server using either low bit rate descriptors (i.e. [6]) or standard 128-SIFT descriptors, according to the needs of the system.

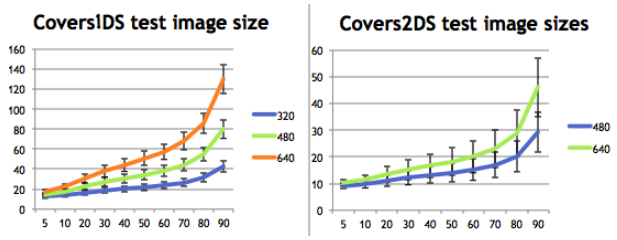

Fig. 4: Sizes on disk of test Images in kilobytes (i.e. for 396 images for Covers1 dataset, 660 images for Covers2 dataset)

4.2.2. MergeF Flag Effect

Comparing Results of Experiments 1 and 2 (i.e. The experiments with $MergeF = False$) to Experiments 3, 4, 5, 6

and 7 (The experiments with $MergeF = True$), we can find that the performance metrics (i.e. Recall, Precision) for Experiments 1 and 2 settles (almost no relative change) starting 20%-20% (Model-Test JPEG quality ratios). While in Experiments 3-7 performance settles starting 30%-30% (Model-Test JPEG quality ratios).

4.2.3. Test Image Resolution Effect

This subsection discusses the effect of changing the max dimension of the test images' resolution, shown in Experiments 3-7 (with $MergeF = True$). Experiments 3, 4 and 5 are on Covers1 dataset, while Experiments 6 and 7 are on Covers2 dataset. In Covers1 dataset, the mean recalls (Recognition Rate) over the 121 sub-experiments in Experiments 3,4 and 5 are 56.93%, 56.35%, and 72.18% respectively. The significant increase in the performance is when the resolution is 640 of test image. In Covers2 dataset (Experiment 6 (480 pixel), 7 (640 pixel)), mean recalls (Recognition Rate) over the 121 cases on Experiments 6 and 7 are 21.07% and 26.86% respectively. This indicates the same behavior relatively on Covers2 dataset. The conclusion here is that the best test max resolution is 640 in both Covers1 and Covers2 datasets and it has a significant influence on the recognition performance.

4.3. Distances and Angles' Effect

From the reported results, it is obvious that the recognition accuracy on Covers2 is worse on average. This is due to distance and angles challenges in Covers2 dataset. In this subsection , The recognition performance is analyzed based on the two datasets in terms of both distances and the angles.

4.3.1. Covers1 Dataset (Experiments 3, 4, and 5)

It is obvious from the results that 640-resolution version of Covers dataset gives the best results. In this subsection, These results of Experiments 3 , 4 and 5 analyze in terms of Distance/ Angle groups . Table 5 compares the recall of the 100% (model image jpeg quality) $\times$ 100% (test image jpeg quality) case for Experiments 3, 4 and 5. The table shows that a significant increase in the recognition of far and medium object while the resolution increases.

100%×100% case	Recall		
Resolution	**320**	**480**	**640**
C - 0 deg	81%	93%	90%
C - 10 deg	86%	93%	90%
C - 30 deg	54%	81%	97%
C - 45 deg	6%	38%	47%
M - 0 deg	47%	88%	95%
M - 10 deg	56%	84%	81%
M - 30 deg	9%	34%	81%
F - 0 deg	0%	20%	61%
F - 10 deg	6%	20%	70%

Table 5: Experiments 3, 4 and 5: Angle / Distance Analysis (100% $\times$ 100% case), where C, M and F denote Close, Moderate distant and Far respectively.

4.3.2. Covers2 Dataset (Experiments 6 and 7)

Table 6 compares the recall of the 100 (Train)$\times$100 (Test) JPEG quality case from Experiments 6 and 7. Similarly, a significant increase in the recognition of far and moderate-distant objects as the resolution increases. Comparing results of Covers1 and Covers2 datasets, it could be concluded, that Distance and Angle complexity does impart the performance more than adding noisy background to the object to be recognized.

100%×100% case	Recall	
Resolution	**480**	**640**
C - 0 deg	90%	93%
C - 10 deg	88%	93%
C - 30 deg	79%	88%
C - 45 deg	54%	65%
C - 60 deg	25%	36%
M - 0 deg	2%	22%
M - 10 deg	6%	22%
M - 30 deg	6%	15%
M - 45 deg	0%	2%
M - 60 deg	0%	2%
F - 0 deg	0%	0%
F - 10 deg	0%	0%
F - 30 deg	0%	0%
F - 45 deg	0%	0%
F - 60 deg	0%	0%

Table 6: Experiments 6 and 7: Angle/Distance Analysis (100% $\times$ 100% case)

5. CONCLUSION

The bottom-line of this study is summarized in the following ten points. (1) The experiments shows logarithmic increase behavior of the recognition performance, while the JPEG quality ratios change in both dimensions (i.e. jpeg quality of model image and test image). The performance settles after certain percentage of JPEG quality (20%-30%). (2) The advantage of lower bit rate descriptors could be combined with the conclusion about JPEG qualities presented in this paper to address the trade-off between the performance and bandwidth limitations. (3) Database size decreases as the JPEG quality increases. However, it settles starting 30% JPEG quality. (4) Distance and Angle complexity is more challenging than adding noisy background to the object to be recognized in the test image. (5) Upsampling increases recognition accuracy with about 10% on average with a penalty of high computational time ($3X$ on average). (6) Image size compared against different quality ratios seems to have exponential growth. This indicates significant drop in the image file size as the JPEG quality decreases, which is beneficial. (7) The best test max-dim resolution is 640 , reflecting a significant effect on recognition performance on both Covers1 and Covers2 datasets. (8) The higher is the test image resolution, the more accurate is the SIFT matching against Distance/Angle challenges. (9) Test images, with angles greater than 45 degree, are poorly recognized even in close pictures using affine matching. (10) The size of object should be around 20% or more of the test image to be fairly recognized.

References

[1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[2] David Nister and Henrik Stewenius, "Scalable recognition with a vocabulary tree," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, Washington, DC, USA, 2006, CVPR '06, pp. 2161–2168, IEEE Computer Society.

[3] B. Girod, V. Chandrasekhar, DM Chen, N.M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, SS Tsai, and R. Vedantham, "Mobile visual search," *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 61–76, 2011.

[4] Yue Wu, Shiyang Lu, Tao Mei, Jian Zhang, and Shipeng Li, "Local visual words coding for low bit rate mobile visual search," in *Proceedings of the 20th ACM international conference on Multimedia*, New York, NY, USA, 2012, MM '12, pp. 989–992, ACM.

[5] Vijay Chandrasekhar, David M. Chen, Zhi Li, Gabriel Takacs, Sam S. Tsai, Radek Grzeszczuk, and Bernd Girod, "Low-rate image retrieval with tree histogram coding," in *Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, ICST, Brussels, Belgium, Belgium, 2009, Mobimedia '09, pp. 7:1–7:7, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[6] Vijay Chandrasekhar, Gabriel Takacs, David M. Chen, Sam S. Tsai, Yuriy Reznik, Radek Grzeszczuk, and Bernd Girod, "Compressed histogram of gradients: A low-bitrate descriptor," *Int. J. Comput. Vision*, vol. 96, no. 3, pp. 384–399, Feb. 2012.

[7] Jie Lin, Ling-Yu Duan, Jie Chen, Rongrong Ji, Siwei Luo, and Wen Gao, "Learning multiple codebooks for low bit rate mobile visual search," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 933 –936.

[8] Hervé Jégou, Romain Tavenard, Matthijs Douze, and Laurent Amsaleg, "Searching in one billion vectors: re-rank with source coding," *CoRR*, vol. abs/1102.3828, 2011.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, june 2005, vol. 1, pp. 886 –893 vol. 1.