Volume xx (200y), Number z, pp. 1–4

Texture Browser: Feature-based Texture Exploration

Supplementary material on the User Evaluation

Xuejiao Luo Leonardo Scandolo Elmar Eisemann

1. Introduction

We conducted a user evaluation to validate the effectiveness and advantages of our method. We wish to perform two evaluations: comparison with alternative methods, and features evaluation. In a first part of our study, we compared against displaying images on a grid (using thumbnails), which is the standard file-system solution. It is the visualization that users typically have the most experience in using. It also requires no metadata (which also holds as is the case for our system). Additionally, we compare to text-based solutions. We also evaluate the four features compared to only using our basic *overview* mode, *multi-scale replication*, *clustering*, *prioritized-tSNE*, and *image-based search*. The first two features aim at general use cases and evaluation of them requires no special settings of the target. Therefore, they were integrated in the first part of the study, acting as additional comparisons. For the latter two features, they aim to improving the retrieval in cases where the number of images is very large, and thus were evaluated separately in the second part of the study.

2. Comparison with alternative methods

Tasks

1. Given a reference texture, users are asked to find it in two ways from the same database: standard grid view, and our texture browser system. The four textures for retrieval are shown in Fig. 1. With our system, users will retrieve the images twice, once starting from the overview without clustering enabled, and later we enforced the cluster view. In this task, we record the times taken for the task for the different procedures. Note that the positions of images in our system changed for each participant, and for each task, since the embedding was rotated by a random angle.
2. Given an abstract descriptive word from the DTD database, users are asked to search for an appropriate texture, which matches this description. In this task, we record the label that the retrieved texture had in the DTD database and compare it to the goal word.

Procedure

1. Provide access to the texture browser software, as well as access to a folder with randomly indexed images for each of the retrieval tasks.
2. Provide a tutorial video of our TextureBrowser tool to familiarize the participants with the GUI and all functionalities of the tool.
3. Let participants carry out the tasks described above, record time, the used functions for each task, and their feedback. Let participants fill in a questionnaire.

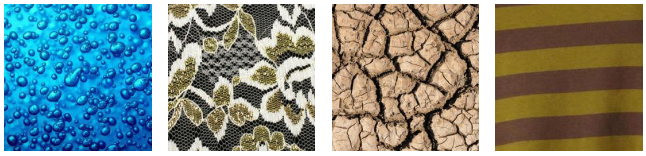

Figure 1: Images used for the retrieval task of methods comparison.

Results and discussion

In total, 16 users participated in our user evaluation. It was their first exposure to our interface and were asked to finish the two tasks listed in the evaluation method above. For each task, retrieval time was recorded and an evaluation questionnaire was filled.

For Task 1, the retrieval times were recorded for the different systems. Table 1 and Table 2 show the detailed timing results, as well as the user satisfaction scores reported by the users in Table 3. The metrics for satisfaction of user interaction is a 5 level Likert scale, where higher means better (specifically, 1: very unsatisfied; 2: unsatisfied; 3: moderate; 4: satisfied; 5: very satisfied).

According to Table 2, the time to retrieve the target textures via the grid interface is substantially reduced via the usage of our tool, either with non-clustered mode or with cluster view mode, with at least a factor of 2, except for user #13 where the reduction is mild. One case worth noticing is that user #15 gave up the retrieval of the last target texture via the grid view after a total search time that exceeded 23 minutes, while succeeding with our solution in less than three minutes (with no cluster view).

On average, the users rated our method higher than the grid view on interaction score. Among all users, only user #9 scored the grid view higher than our system, and only for the cluster view

User	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Grid																
Texture 1	0:10	0:41	0:58	0:23	0:44	0:48	0:34	0:28	0:35	1:30	0:27	0:48	1:16	0:46	0:18	4:02
Texture 2	5:07	2:06	1:01	0:43	0:08	5:40	1:10	1:55	2:21	0:51	4:44	0:29	1:00	1:05	3:23	2:15
Texture 3	0:47	1:17	1:23	0:17	1:04	2:00	0:27	1:37	2:02	1:00	0:56	0:30	0:50	0:27	1:48	1:33
Texture 4	0:39	5:18	4:11	0:48	1:46	1:26	2:00	1:46	0:58	1:58	1:02	2:20	1:10	1:26	17:40	3:00
Ours																
Texture 1	0:05	0:17	0:11	0:11	0:18	0:15	0:08	0:10	0:06	0:26	0:13	0:13	1:00	0:08	0:20	0:17
Texture 2	0:35	1:24	1:40	0:19	0:38	1:00	1:26	0:10	0:25	0:39	0:59	0:11	0:46	0:27	1:06	0:38
Texture 3	0:20	0:43	0:08	0:14	0:05	0:20	0:14	0:24	0:20	0:23	0:20	0:20	1:10	0:40	0:43	0:26
Texture 4	2:26	2:41	0:15	0:20	1:03	0:03	0:13	0:29	0:17	1:09	0:13	0:30	0:40	0:14	0:14	0:58
Ours (cluster view)																
Texture 1	0:30	0:32	0:10	0:23	0:25	0:46	0:20	0:10	0:21	1:03	0:09	0:40	0:45	0:41	0:43	0:21
Texture 2	0:46	0:51	0:32	0:16	0:31	0:40	0:36	0:15	0:30	0:47	1:40	0:31	0:25	0:22	1:26	0:57
Texture 3	0:23	0:12	0:46	0:18	0:27	0:56	0:12	0:35	0:18	0:27	0:39	0:31	0:46	0:27	0:57	0:31
Texture 4	0:19	0:59	0:07	0:13	0:22	0:22	0:21	0:50	0:27	1:10	1:22	0:22	0:50	0:26	0:42	0:25

Table 1: Task 1 retrieval times for individual textures

User	Grid	Ours	Ours (clustering)
1	06:43	03:26	01:58
2	09:22	05:05	02:34
3	07:33	02:22	01:35
4	02:08	01:04	01:10
5	03:42	02:04	01:45
6	09:54	01:38	02:44
7	04:11	02:01	01:29
8	05:46	01:13	01:50
9	05:56	01:08	01:36
10	05:19	02:37	03:27
11	07:09	01:45	03:50
12	04:07	01:14	02:04
13	04:16	03:36	02:46
14	03:44	01:29	01:56
15	$\geq$23:09	02:23	03:48
16	10:50	02:19	02:14
Average	07:07	02:13	02:18

Table 2: Total user timings for task 1

User	Grid	Ours	Ours (clustering)
1	3	4	4.5
2	2	4	4
3	1	3	5
4	2	5	4
5	2	4	3
6	1	4	3
7	2	5	4
8	1	5	3
9	4	5	3
10	2	4	3
11	3	5	4
12	2	4	4
13	3	3	4
14	2	5	4
15	1	4	4
16	1	4	5
Average	2	4.25	3.84

Table 3: User interaction scores

mode, despite having a shorter retrieval time in our system with the cluster view. Overall, the users preferred their interaction without the cluster view (avg. score 4.25) slightly over the cluster view (avg. score 3.84).

From open feedback the users provided on the questionnaire, the main reason of preference for the unclustered exploration is that it feels more intuitive for them. This is reasonable since the cluster view requires more involved control (i.e., defining an appropriate number of clusters). Furthermore, some users noted that the resulting clusters did not group textures according to their expectation, and choosing the wrong cluster to explore would not allow them to find the target texture. Nevertheless, some users noted that the cluster view is better suited for retrieval of textures with very well defined features, such that the corresponding cluster is well defined and easier to find, especially when the overall number of textures is large. With a more sophisticated clustering method, this situation can be alleviated. Finally, some users also mentioned that they believe that training would improve their performance.

Multi-scale We also evaluated the effectiveness of our multiscale replication. The first target texture among the given four was present twice in the embedding, via the embed position of the feature vector for the original version and one of the blurred versions. Out of 16 participants, 12 obtained the target based on the position corresponding to the blurred version, in an average 27 seconds, while the rest obtained the target in its original embedding position, in an average 38 seconds. This validates our hypothesis that different users first

User	DTD Word	DTD Word
	Keyword: Porous	Keyword: Interlaced
1	 Porous	 Zigzagged
2	 Honeycombed	 Grid
3	 Scaly	 Woven
4	 Porous	 Crosshatched
5	 Porous	 Braided
6	 Porous	 Braided
7	 Fibrous	 Braided
8	 Bubbly	 Braided
9	 Pitted	 Woven
10	 Porous	 Woven
11	 Flecked	 Woven
12	 Pitted	 Braided
13	 Porous	 Braided
14	 Porous	 Knitted
15	 Porous	 Interlaced
16	 Dotted	 Banded

Table 4: Images selected by users in task 2

search for textures in groups according to features of different scales, and our multiscale replication allows them to find the texture in either case. On average, users were inclined to retrieve this image via the features of its blurred version, and the retrieval time in that case is shorter. We believe that this is because the large-scale features are easier to identify starting from an unzoomed view, as is the case in our tool.

Text-based retrieval In Task 2, each user retrieved two textures via our interface according to their understanding of the two descriptive words they were given. These words are taken from the Describable Textures Dataset (DTD), which contains 1519 images, and groups them according to 47 words (terms/categories) inspired from human perception. The retrieved textures for each participant are shown in Table 4. The words provided to the users were *porous* and *interlaced*.

From the results shown in Table 4, we can see that the interpretation of a specific adjective differs from person to person. For the first word, *porous*, half of the users obtained a texture whose describing word from DTD is identical to the given word. For the other half, the retrieved images have a common feature of small holes, which could roughly match the expectation of the word *porous*.

For the second word, *interlaced*, only one user (#15) retrieved a texture that is labeled with that word in DTD. Most of the users retrieved textures that can be described with the words *woven* and *braided*. These words can clearly describe texture features that are very similar, and we believe most non-expert users of such a word-based classification system would struggle to differentiate among them, as shown by our results.

3. Prioritized t-SNE and image-based search tool evaluation

The prioritized-tSNE and image-based search features were evaluated in a second part of the study. We divided the participants into two groups (group A and group B) to evaluate each feature independently, group A without using any feature and group B using the specified features. Since these two features aim at solving retrieval in a large and crowded database, the texture database used here contains 5824 textures.

Tasks

1. **Prioritized-tSNE:** Given a reference texture, users are asked to find the identical texture via our interface. For group A, users can browse around freely with only the overview panel of our interface, but without using any of the advanced features (prioritized-tSNE, image-based search, or clustering). For group B, users first try to select a texture similar to the target texture from the overview panel as an input to the prioritized-tSNE tool, and assign it with a large weight to spread out the crowd where the target texture may belong. After that, they are suggested to browse around the highlighted vicinity for further retrieval. The textures for retrieval are shown as the two in the left in Fig. 2.
2. **Image-based search:** Given a reference texture, users are asked to find the identical texture using our interface. For group A, users can again browse around freely with only the overview panel of our interface, but without using any of the advanced features. For group B, users first use our image-based tool to locate a similar texture in the database, after which they can freely browse until they locate the reference image. We provide two textures similar to the reference texture as as input to the image-based retrieval, and users are also allowed to draw sketches as well. The reference textures for retrieval are shown as the two in the right in Fig. 2.

Procedure

Figure 2: Images used for the retrieval tasks of features evaluation.

1. Provide access to our texture browser software.
2. Provide a tutorial video of how to use overview mode of the interface, i.e., zoom in/out, translate, select images and tiling in a grid, to group A, and a tutorial video of how to use both overview mode and the prioritized-tSNE and image-based search tools to group B.
3. Let participants carry out the tasks described above, record time, and their feedback (optional). Let participants fill in a questionnaire.

Results and discussion

In total 16 users were randomly devided into two groups, each with 8 users. The recorded retrieval time for each texture is shown in Table 5 and Table 6.

Prioritized t-SNE According to Table 5, in the prioritized-tSNE evaluation, the average time of group A (using only the overview mode) is more than two times that of group B (using the prioritized-tSNE tool). This suggests that by using our prioritized-tSNE tool, the retrieval can be substantially faster. There is an outlier, user #13, who was misled by a similar texture, an image of cracked glass, and spent a long time in the wrong region according to the feedback. This suggests that when using our prioritized-tSNE tool, it is important for the user to have an impression of the overall distribution first, such that an appropriate image can be selected as input.

From the free feedback in the questionnaire, some users from the group A suggested that a tool that can spread the overlapped crowd out would be helpful. This also reflects the usefulness of the prioritized-tSNE tool.

Image-based search In the evaluation of the image-based search tool, as shown in Table 6, the average time to finish the task of group A (using only the overview mode) is at least 4 times that of group B (using the image-based search tool). In group A, two users (#2 and #4) gave up the task after 5 minutes and 3 minutes, respectively. From their feedback, we learned that user #2 switched the retrieval among several possible regions but was still not able to locate the target. User #4 gave up retrieval due to overlapping textures and suggested that a tool to spread the crowded region out would be useful. In group B, almost all the users succeeded in finishing the task using less time than those in group A, except for user #10 for whom the retrieval was stopped at 5 minutes and the reason was similar to that of user #4. This can probably be solved by the use of our prioritized-tSNE tool. Overall, thanks to the fast identification of the target region in the large database when using the tested tool, the retrieval time was substantially shorter.

Group	User	Texture 1	Texture 2	Total
A	1	04:02	06:54	10:56
A	2	02:38	01:36	04:14
A	3	01:15	02:46	04:01
A	4	04:04	00:43	04:47
A	5	01:15	02:05	03:20
A	6	01:10	03:30	04:40
A	7	03:52	04:30	08:22
A	8	04:56	01:57	06:53
A	**Average**	02:54	03:00	05:54
B	9	01:06	00:53	01:59
B	10	00:52	00:54	01:46
B	11	00:43	01:01	01:44
B	12	00:31	00:52	01:23
B	13	10:34	00:32	11:06
B	14	01:09	00:40	01:49
B	15	01:12	01:07	02:19
B	16	01:03	00:42	01:45
B	**Average**	02:08	00:50	02:58

Table 5: Retrieval times for individual textures in feature “prioritized-tSNE” evaluation

Group	User	Texture 3	Texture 4	Total
A	1	13:08	16:00	29:08
A	2	≥05:00	≥05:00	≥10:00
A	3	07:06	05:28	12:34
A	4	≥03:00	01:23	≥04:23
A	5	04:14	03:58	08:12
A	6	10:00	01:30	11:30
A	7	00:47	02:42	03:29
A	8	04:57	02:32	07:29
A	**Average**	06:01	04:49	10:50
B	9	00:20	01:33	01:53
B	10	≥05:00	00:30	≥05:30
B	11	01:02	00:58	02:00
B	12	00:43	02:07	02:50
B	13	01:05	02:03	03:08
B	14	00:40	00:30	01:10
B	15	00:43	00:25	01:08
B	16	01:21	00:43	02:04
B	**Average**	01:21	01:06	02:27

Table 6: Retrieval times for individual textures in feature “sketching” evaluation