FAVORED WORKFLOWS IN LIGHT FIELD EDITING

Belen Masia, Adrian Jarabo and Diego Gutierrez
Universidad de Zaragoza
{bmasia|ajarabo|diegog}@unizar.es

ABSTRACT

Light fields —4D representations of a scene coding luminance as a function of both position and angle— are becoming increasingly easy to capture. A number of handheld light field cameras, such as those by Lytro™, Raytrix™, or Pelican Imaging™, have already entered the consumer market and are bringing light field photography to the wide public. This gives rise to a new need: Interfaces and tools for light field editing. In this work we provide a detailed description of subjects' performance and preferences for a number of possible editing tasks. We believe these insights can aid researchers and designers in their creation of new light field editing tools and interfaces.

KEYWORDS

Light fields; editing interfaces; user study; editing workflow.

1. INTRODUCTION

Light fields are four-dimensional representations of a scene, where the two extra dimensions code angular information about the scene being captured. This allows to create effects such as small parallax shifts, synthetic refocusing after capture, or even scene reconstruction, leveraging this additional wealth of information. With the introduction in the consumer market of light field cameras (such as Lytro (2014), Raytrix (2014) or the new Pelican camera (2014)), they are becoming an increasingly popular alternative to traditional 2D images.

However, editing light fields remains an open problem: While editing such traditional 2D photographs is a well-understood process with an established workflow (using popular software like Photoshop™), it is not clear how to edit the four-dimensional information stored in a light field. Jarabo et al. (2014) recently presented the first comprehensive study on the topic: They evaluated a set of basic tools on two light field interface paradigms (parallax-based and focus-based), asking participants to perform several edits on both synthetic and real light field data available online (such as changing the color of an object, painting on its surface, altering exposure, etc).

In their work, they provided valuable insights on aspects such as choice of tools or interface preference, which should help guide future (and needed) light field editing applications. Here, we provide additional data and insights not present in that work, focusing on another important aspect in editing: We focus on subjects' workflows, and, for a number of typical scenarios (editing of surfaces, editing in free space, occlusion handling, and editing of complex geometries) we describe in detail the trends found. This provides a more comprehensive description of subjects' choices and preferences with regard to light field editing interfaces which can be used as a basis for future work in the field, both in interface and tool development.

2. RELATED WORK

There is a body of work in tools for light field editing which implicitly use certain interfaces for performing the edits. In their early work, Seitz and Kutulakos (2002) perform edits such as painting and scissoring by creating a voxel-based representation of the light field to propagate the edits. Zhang et al. (2002) and Wang et al. (2005) perform morphing of two light fields, for which users need to indicate corresponding polygons between them to guide the process. User interaction is also required in the Pop-Up Light Field work by Shum et al. (2004), in which users mark the silhouettes of objects in multiple views, which is then used to separate the light field into depth layers. Jarabo et al. (2011) devise a system for propagation of edits between views based on pixel affinity measured in a multidimensional space. Also somehow related are the works on editing of stereo pairs, such as that of Stavrakis and Gelautz (2004), or Lo et al. (2010). All these works focus on the editing tools, but not on the interaction required by the user to perform them, as we do in this work.

In contrast to editing tools, light field editing interfaces remain largely unexplored. Here, we build on the recent work of Jarabo et al. (2014) in this realm. While the focus of that work is on whether certain interfaces allow for editing of light fields, which tools are appropriate, and whether current reconstruction methods are suitable for editing, in this paper we describe subjects' workflows and preferences focusing on the type of editing scenario, namely editing of surfaces, editing in free space, occlusion handling, and editing of complex geometries.

3. LIGHT FIELD EDITING INTERFACES AND TOOLS

Interaction paradigms: The majority of previous work which has explored editing of a light field has relied on correspondences between the views for specification of the edits to perform, e.g. (Seitz and Kutulakos 2002; Stavrakis and Gelautz 2004; Lo et al. 2010). This amounts to using parallax as a depth cue to specify the depth at which an edit should be placed. We term this approach a *multiview* interaction paradigm. A completely different paradigm based on Isaksen et al.'s light field reparameterization (2000) was explored by Davis et al. (2012); depth at which the editing will be performed is specified by a plane of focus, while the rest of the light field is blurred accordingly (wide aperture rendering). This amounts to using blur as a depth cue to specify an edit's position in depth, and we term this approach *focus*. Thus, both paradigms differ in how the *depth* at which the edit is to be placed is specified. These two are the paradigms explored in the tests.

In the following we describe the procedure to perform an edit in the light field with each of the basic interaction paradigms we use, multiview and focus. Note that we work with *point-based, local edits* (strokes). Our assumption is that a local point-based interaction lies at the basis of the majority of editing processes and tools (from the use of a brush or the eraser to more complex editing tools like selection or global filters). To place a stroke in *multiview*, the user first draws it in one view; the epipolar lines of the stroke then become visible, and by switching between views and observing the resulting parallax of the stroke the user can infer its depth and place it in the desired position in the light field (Figure 1, *left*, (a)). In *focus*, first, the plane of focus is selected, and then a stroke drawn, which will lie, in the light field, in the in-focus plane selected (Figure Figure 1, *left*, (b)). Essentially this means that in multiview the stroke is first drawn, and then its position adjusted, while in focus the depth is selected first, and then the stroke drawn.

Interfaces: The two first interfaces we study are derived from the interaction paradigms described above (Section 3.1), and are thus the *Multiview* (**M**) interface and the *Focus* (**F**) interface. In addition to these, there are two more interfaces, described next.

Recently, research in scene reconstruction from light fields has undergone significant progress (Wanner and Goldlücke 2012; Kim et al. 2013), yielding depth maps of a scene computed from a light field of it. Having depth information together with the input light field may alter the process preferred by users for editing a light field. As such, in the Multiview and Focus interfaces above, we include the option to use depth, yielding the *Focus With Depth* (**FD**) and the *Multiview With Depth* (**MD**) interfaces. In both, this means that the edits (the strokes) drawn will snap to the nearest surface below them. Therefore, the main difference between these two interfaces will be in the visualization of the light field and the performed edits: While in Multiview With Depth these are visualized switching the point of view, in Focus With Depth this is done shifting the in-focus plane.

Editing tools: The interfaces tested feature, in the first place, a set of basic tools which includes the following: Painting brush, erasing brush, dodging brush, burning brush, and pasting of pre-loaded billboard objects. They further include a *depth selection* tool, which allows the user to pick (mouse-clicking) a certain depth, and specify a depth range around it (using a slider to set the threshold). This results in a selection of only that range in depth, outside of which editing cannot be performed (like a mask in depth). Similar in spirit to this tool, a *color selection* tool is also included. With it, the user selects a color, and a threshold, creating a selection mask based in color. Finally, a *visual aid* tool is added, with which the areas of the light fields which are selected at a certain time are highlighted (by adding a semi-transparent checkerboard pattern to the non-selected areas).

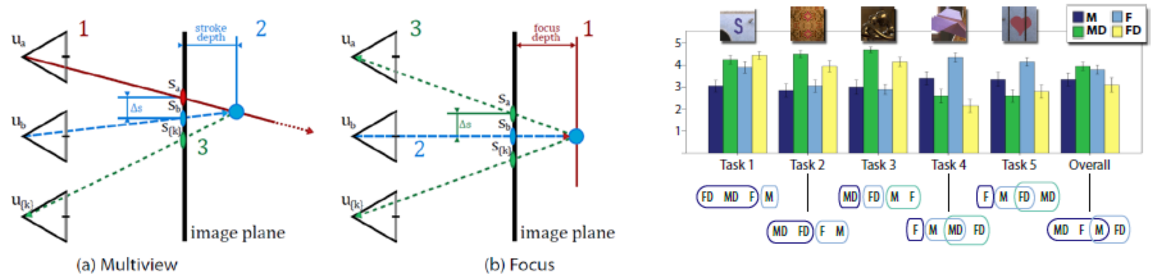

Figure 1. **Left**. Procedure to place a stroke in the light field in both paradigms tested. (a) Multiview: The stroke s_a is first drawn in a view u_a (1), and then its depth adjusted in another view u_b by displacing the stroke along the epipolar line (2), once this depth is chosen, the stroke is fixed and propagated to other views $u_{\{k\}}$. (b) Focus: First the depth is selected by adjusting the in-focus plane (1), and a stroke s_b is placed in the central view u_b (2), which is then propagated to the other views $u_{\{k\}}$ (3). **Right**. Rating results for Experiment 1. Mean ratings for the five directed tasks in Experiment 1, and overall rating per interface. Results of the pairwise comparisons between interfaces are also shown, indicating which differences are significant (interfaces in the same group are not significantly different). *Source:* (Jarabo et al. 2014).

4. EXPERIMENTAL DESIGN

Two experiments are performed, one with synthetic scenarios, and 20 subjects, and the second one with real scenarios, and 10 subjects. In both of them the subjects are asked to perform several tasks using editing interfaces, and both objective and subjective data is collected. Objective data comprises timing data (such as time to completion) and measures of error in depth (when available, as explained next). Subjective data includes ratings and rankings by users collected via questionnaires and free form comments.

4.1 Synthetic Scenarios (Experiment 1)

Stimuli: We use three synthetic light fields for the different tasks, which can be seen in Figure 2. They are chosen to cover a variety of scenes, depth, and reflectance complexities. For each scene 17x17 views are rendered, and they are presented at a spatial resolution of 600x600 to ensure real time performance of the interfaces. Since they are synthetic scenes, the actual ground truth geometry and depth are known, which can be used to compute the error in depth in which subjects incur when performing the different tasks.

Draw on wall | Change color of vase | Edit specular highlights | Add airplane in free space | Draw behind the railing

Figure 2. Tasks S1 to S5. Target image and description for Tasks S1 to S5 used in Experiment 1.

Tasks: The first experiment contains seven tasks [S1..S7], which are performed in fixed order. The first five tasks are directed tasks, that is, a target image and detailed explanation of the editing to perform is given to subjects. The final two tasks are open tasks, where there is no target image; instead, some photographs are given for inspiration, but the subject is free to perform any editing she wants, given she has already become familiar with the different tools and interfaces. Tasks are chosen so that subjects perform a variety of edits, including editing planar (S1) and curved surfaces (S2), highlights (S3), placing objects in free space (S4), or dealing with occlusions (S5). Figure 2 details the tasks, with the target images given to subjects, and short description of each. Tasks S6 and S7 will be deliberately left out of the discussion in this work, since the fact that they are open creates very large variability between subjects.

Procedure: Twenty subjects (6 female, 14 male) took part in the experiment. Most of them did not have previous knowledge of light fields, but all had knowledge on image editing or 3D modeling software. Each subject was presented with the four interfaces, in random order to avoid possible learning effects, and for each interface she had to perform the seven tasks in fixed order, S1 to S7. Time per task was limited to five minutes in directed tasks, and twelve in open ones. After each interface and at the end of the experiment subjects had to fill in questionnaires regarding their preferences, and then a debriefing session was carried out. Subjects also underwent a training session (ca. one hour long) in which they learnt about light fields and how to use the different interfaces. Note that in this experiment, the depth selection, color selection and visual aid tools were not included or available to subjects.

4.2 Real Scenarios (Experiment 2)

In Experiment 2, a hybrid interface combining the different interfaces above that allows users to switch between them, and including the three new tools (depth selection, color selection and visual aid tools) was used by the subjects. These additions and modifications are based on the insights acquired via the first experiment. The goal here was to test whether, with this new interface and tools, subjects could perform *common*, everyday editing tasks on *real* light fields satisfactorily. Again, we describe here the methodology, while the results are analyzed in Section 5.

Stimuli: We use eight real, captured light fields for the different tasks, which can be seen in Figure 3. They are once again chosen to cover a variety of scenes, depth, and reflectance complexities. Three of them have 9x9 views and are captured with a Lytro™ light field camera (Lytro 2014), and their depth obtained with the Lytro SDK; four of them (17x1 views) are captured with a camera gantry, and thus have a wider baseline, and their depth is obtained with the algorithms from Kim et al. (2013); a final scene (9x9 views) is captured with a Raytrix™ light field camera (Raytrix 2014), and its depth reconstructed with the algorithm by Wanner and Goldlücke (2012). In this case, since we work with real light fields, we do not have ground truth depth information; however, the goal is to test if the interfaces enable users to perform common edits satisfactorily, regardless of the actual accuracy we are interested in subjective opinion.

Tasks: This experiment contains ten directed tasks [R1..R10], in which a target image and detailed explanation of the editing to perform is given to subjects. Again, tasks are chosen to cover a variety of edits. Figure 3 shows the task, including for each one both the input central view and the target images given to subjects.

Procedure: Ten subjects (4 female, 6 male) took part in the experiment. Half of them did not have previous knowledge of light fields, and all had knowledge on image editing or 3D modeling software. Each subject was presented with the ten tasks in random order to avoid possible learning effects. To carry out the tasks, she could choose freely between a multiview and a focus paradigm, and between using or not depth information. Time per task was limited to ten minutes. After each task subjects filled in a short questionnaire with several ratings (e.g., difficulty of task, or similarity to target). Another questionnaire was presented at the end of all ten tasks, and a debriefing session also took place. Additionally, timings (times of use of the interaction paradigms, of depth information, and of the different tools) were collected to analyze the different workflows. Before proceeding with the tasks, subjects underwent a training session.

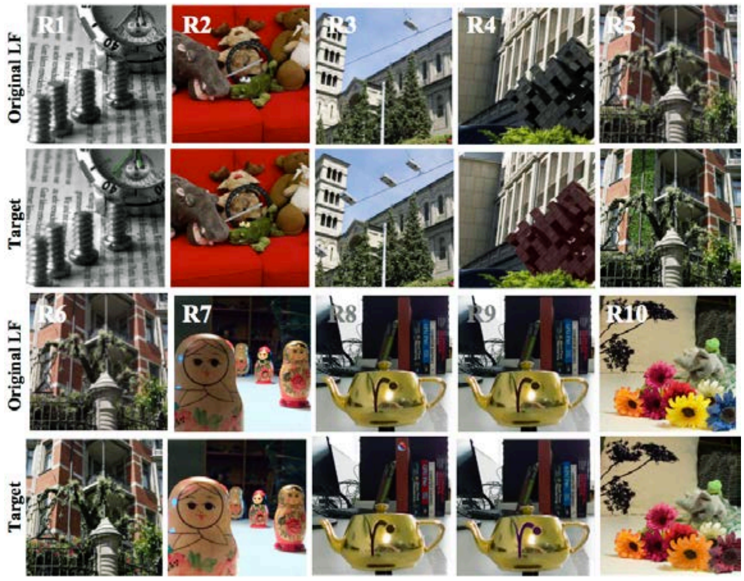

Figure 3. Tasks R1 to R10. Input image and target image for Tasks R1 to R10 used in Experiment 2. R1 is based on Wanner and Goldlucke (2012), R2 to R6 are based on Kim et al. (2013), and R7 to R10 are from Lytro.

5. WORKFLOW ANALYSIS

Data collection from the experiments described above yields an immense quantity of both subjective and objective information which provides insights on a variety of different aspects of light field editing: Usage of different tools, preferred interfaces, variability of preferences with the task to perform, workflows, etc. While previous work has focused on how suitable different interfaces and tools are, and whether an interface can allow for satisfactory editing of light fields with current depth reconstruction methods (Jarabo et al. 2014), here we focus on subjects' workflows, that is, we look for underlying patterns in subjects' actions and their preferences for different generic tasks.

5.1 Synthetic Scenarios (Experiment 1)

Analysis of the data for time to completion, error in depth (measured as the L1 norm averaged across views), and ratings and rankings on interface preference provided by users yielded three distinct clusters, roughly corresponding to three task categories: editing of surfaces (planar or curved), editing in free space, and occlusion handling. Data analysis is performed using repeated measures ANOVA for error, timings, and ratings, and Kruskal-Wallis for rankings (Cunningham and Wallraven 2011).

Editing of surfaces: Tasks S1 to S3 allow us to draw insights on surface editing. In terms of error, in these tasks error in depth for interfaces with depth (MD and FD) is zero, since strokes snap to the surface below them (Figure 4, *left*). In a consequent manner, realizing the task with these interfaces took less time (Figure 4, *right*). For interfaces without depth, M yielded a higher error ($p \leq 0.018$) than F ($p \leq 0.018$), showing that users found it more difficult to locate an edit in depth with M. This is also reflected in the timings, in which M is significantly slower ($p \leq 0.008$) in most cases. Task S1 is an exception, possibly because of its simplicity. Ratings show a clear preference for interfaces with depth for editing on surfaces, as expected. The correlation of rankings with the results for ratings is extremely high, showing that subjects have clear preferences. In summary, interfaces with depth are the clear choice for editing surfaces, as expected. Further, a slight preference for multiview over focus is hinted (see Figure 1, *right*), but the difference is not significant

Editing in free space: Task S4 deals with positioning in free space. In this case, interfaces without depth yield lower error than those with depth, although the difference between interaction paradigms is not significant (Figure 4, *left*). Even though errors are high, times to completion is very low (Figure 4, *right*). In interfaces with depth, this can be due to the fact that users realize that they will not be able to correctly place it in depth (recall that the edit will snap to the surface right below) and give up. This hypothesis seems supported by the low ratings these two interfaces receive in this task (Figure 1, *right*). In general, results suggest that users struggle to correctly judge depth in free space, and feedback from users confirmed this fact, and pointed out that F is chosen because of the clear feedback it provides with respect to the position of the plane being edited. In accordance with this, F takes the least time, even though the difference is only significant with respect to M ($p \leq 0.008$). Thus, F is the interface of choice among the four tested for editing in free space.

Figure 4. **Left**: Mean error in depth for the five directed tasks in Experiment 1, and results of the pairwise comparisons between interfaces *Source:* (Jarabo et al. 2014). **Right**: Same for time to completion.

Occlusion handling: Task S5 implies handling occlusions. For this, F yields both the lowest error and time to completion (although not significantly different from FD in the latter and from MD in the former) (Figure 4). Clearly, M and MD are not useful interfaces in this regard, since M yields the largest error, and MD the highest time to completion, to the point that some subjects (35%) did not complete the task in the given time. Note that in this experiment, subjects need to erase to handle occlusions, which can be time consuming. Ratings confirm the superiority of F in this task (see Figure 1, *right*).

In summary, this experiment has shown the large influence of the task to perform in the interface of choice, with, essentially, MD and FD being ahead for on-surface editing, and F the interface of choice for free-space editing and occlusion handling. In the following, we will see if these findings hold in real light fields, with new tools and subjects being freely allowed to choose the interaction paradigm and whether depth information should be used (essentially, being able to switch between the four previous interfaces).

5.2 Real Scenarios (Experiment 2)

Again, we cluster the tasks in categories; these are the same as the above, with an additional category for editing of objects of intricate geometry, for which subjects follow a different workflow than when editing simpler planar or curved surfaces. As mentioned above, our focus here will be on workflows, and we will look at times of use of the different interfaces and tools as an indicator of preferences. We also collect data on subjective preferences. We describe here the main findings.

Editing of surfaces: Tasks R1 and R8 involved editing planar surfaces, while R2, R7 and R9 require editing curved surfaces (see Figure 3 for reference). Among the latter, R2 and R9 are very similar in nature, since they both involve changing small details, while R7 requires editing a large curved surface. In all of these tasks, the use of depth information is largely favored, which matches our results in Experiment 1. There is no clear preference between interaction paradigms (focus or multiview), although there is a slight trend towards focus; in the debriefing interviews, subjects reported that focus offered a very strong and easy-to-interpret cue for visualization of the active area. Regarding the tools used, Tasks R2 and R9 favor the use of the color selection tool, possibly because, as mentioned, the areas requiring editing are small, similar in color, and without a distinct depth with respect to their surrounding areas. The rest favor the use of the depth selection tool. A sample editing workflow (for Task R1) is shown in Figure 5.

Editing in free space: Task R3 requires editing in free space. In this case, again, results are consistent with Experiment 1: Depth information is scarcely used, if at all. However, while, before, F was preferred, we observe here a trend towards M, as shown in the workflow for Task R3 included in Figure 5. This is possibly due to the absence of high frequency information in the area to edit, which causes the blur of the focus interface to provide little or no depth cues.

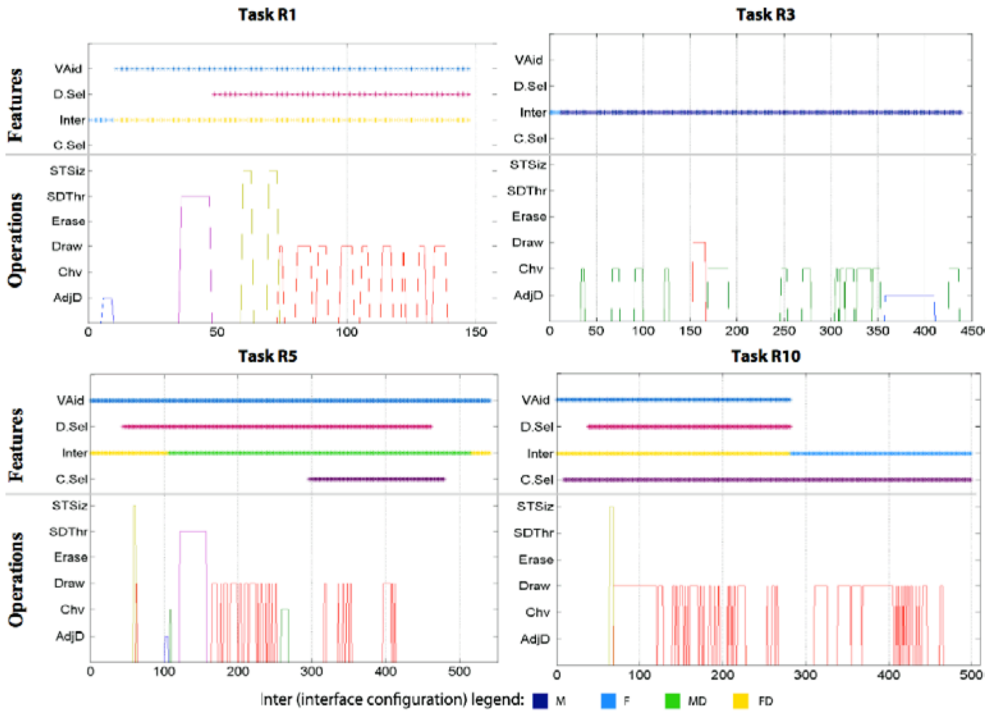

Figure 5. Sample workflows for Tasks R1, R3, R5 and R10, one pertaining to each of the four categories or use case scenarios (editing surfaces, editing in free space, handling occlusions, and editing complex geometries). They indicate the tools and interfaces used by the subject along time, shown in the abscissa (in seconds). The six bottom marks in the y-axis correspond to different operations (from bottom to top: adjust depth, change view, draw, erase, set depth threshold, and set tool size). The top four marks indicate whether that feature was activated or not at each time instant, and correspond, from bottom to top, to color selection tool (*C.Sel*), interface (*Inter*), depth selection tool (*D.Sel*) and visual aid tool (*VAid*). Note that the *Inter* has features different colors specifying the interface being used. These workflows for all subjects and tasks can be found at http://giga.cps.unizar.es/~ajarabo/pubs/lfeiSIG14/.

Occlusion handling: In this experiment, Tasks R5 and R6 require dealing with occlusions. Here, the introduction of the depth and color selection tools causes a change with respect to results obtained in Experiment 1. While in the first experiment, there was a large amount of erasing to deal with the occlusions, the introduction of the depth selection tool, largely used in both R5 and R6, reduces the need to erase to a minimum (see Figure 5, Task R5, for a sample editing workflow in that task). Surprisingly, there is little difference between the use of MD and FD, revealing that as long as depth information and related tools are present, the interaction paradigm is less relevant for these tasks. The color selection tool is fairly used in Task R5 to avoid the pipe, which is hard to disambiguate from the rest of the wall in the depth dimension.

Editing of complex geometries: Tasks R4 and R10 clearly show the need for the color and depth selection tools. When intricate geometries are present, these are extensively used. The nature of the scene determines which one is used: In the case of R4, 9 out of 10 subjects used the depth selection tool to complete the task, while in the case of R10, 9 out of 10 used the color selection tool, as shown in the sample editing workflow for Task R10 shown in Figure 5. The majority of the subjects used depth information throughout the tasks; however, differences between the time of use of FD and MD were not significant.

In summary, the second experiment confirms the findings of Experiment 1 in most aspects, with a clear exception in occlusion handling, which is now easily dealt with thanks to the new tools. Similarly, handling of intricate geometries is possible thanks to these tools. We also observe that depth information is almost always required, while the differences between the interface paradigms (multiview and focus) become less significant.

6. CONCLUSIONS

We have tested a set of interfaces and tools for light field editing, and showed that they allow users to satisfactorily edit light fields and perform common, everyday tasks, such as those in Experiment 2, on real, captured, light fields. While previous work has focused on the feasibility of using the light field editing interfaces, tools, and depth reconstruction methods, we present here the main findings in terms of workflow and preferences for the different scenarios (free space, planar surfaces, occlusions, etc.) which a user may encounter. Light field editing is still in its infancy, and we believe this work contributes to the foundations and to providing a solid basis for the development of future tools and interfaces, both for researchers and UI designers. In order to further help future research, we have the code of the interface and the full datasets available on-line at http://giga.cps.unizar.es/~ajarabo/pubs/lfeiSIG14/ .

ACKNOWLEDGEMENTS

We want to thank Adrien Bousseau, Fabio Pellacini, the participants of the experiments (in particular Sara Lopez, Patrick Moosebrugger, Carlos Aliaga and Jose I. Echevarria), Chia-Kai Liang for his help with the Lytro SDK, and the members of the Graphics and Imaging Lab. We also thank Guillermo Leal Llaguno for the San Miguel scene, Emmanuelle Chapoulie and the AIM@SHAPE project for vase, and Infinite Realities for the head data. This research has been partially funded by the EC through projects VERVE (ICT) and GOLEM (Marie Curie IAPP), and the Diputación General de Aragón (project TAMA). Belen Masia has additionally been supported by an Nvidia Graduate Fellowship.

REFERENCES

Cunningham, D., and Wallraven, C., 2011. *Experimental Design: From User Studies to Psychophysics.* A. K. Peters.

Davis, A. et al, 2012. Unstructured Light Fields. *In Computer Graphics Forum*, Vol. 31, No. 2, pp. 305-314.

Isaksen, A. et al, 2000. Dynamically Reparameterized Light Fields. *In ACM Trans. on Graphics*, V. 22, No. 4, 297-306.

Jarabo, A. et al, 2011. Efficient Propagation of Light Field Edits. *In Proc. of SIACG*, Faro, Portugal, pp. 75-80.

Jarabo, A. et al, 2014. How Do People Edit Light Fields? *In ACM Transactions on Graphics*, Vol. 33, No. 4.

Kim, C. et al, 2013. Scene Reconstruction from High Spatio-Angular Resolution Light Fields. *In ACM Transactions on Graphics*, Vol. 32, No. 4, pp. 73:1-12.

Lytro 2014. The Lytro^{TM} Camera. www.lytro.com

Lo, W.-Y. et al, 2010. Stereoscopic 3D Copy & Paste. *In ACM Transactions on Graphics*, Vol. 29, No. 6, pp. 147:1-10.

Pelican 2014. Pelican Imaging Light Field Camera. http://pelicanimaging.com/technology/camera.html

Raytrix 2014. The Raytrix^{TM} Camera. www.raytrix.de

Seitz, S. M. and Kutulakos, K., 2002. Plenoptic Image Editing. *In Intnl. J. on Comp. Vision*, Vol. 48, No. 2, pp. 115-129.

Shum, H.-Y. et al, 2004. Pop-Up Light Field: An Interactive, Image-Based Modeling and Rendering System. *In ACM Trans. on Graphics*, Vol. 23, No. 2, pp. 143-162.

Stavrakis, E. and Gelautz, M., 2004. Image-based Stereoscopic Painterly Rendering. *In Proc. of EGSR 2004*, pp. 53-60.

Wang, L. et al, 2005. Light Field Morphing Using 2D Features. *IEEE Trans. Vis. Comp. Graph.*, V. 11, No. 1, pp. 25-34.

Wanner, S. and Goldlücke, B. 2012. Globally-consistent Depth Labeling of 4D Light Fields. *In Proc. of CVPR'12*, 41-48.

Zhang, Z. et al, 2002. Feature-Based Light Field Morphing. *In ACM Trans. on Graph.*, Vol. 21, No. 3, pp. 457-464