Image-Based Information Filtering to Compare and Select Items

The editorial version of the paper can be found at:
Filippo Hu Z., Mauro N., Ardissono L., Image-Based Information Filtering to Compare and Select Items (2023),
22nd IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2023, pp. 1 - 8,
DOI: 10.1109/WI-IAT59888.2023.00007

Image-based Information Filtering to Compare and Select Items

1st Zhongli Filippo Hu
Computer Science Department
University of Torino
Torino, Italy
zhonglifilippo.hu@unito.it

2nd Noemi Mauro
Computer Science Department
University of Torino
Torino, Italy
noemi.mauro@unito.it

3rd Liliana Ardissono
Computer Science Department
University of Torino
Torino, Italy
liliana.ardissono@unito.it

***Abstract*—Current recommender systems overlook the role of images in conveying information about items, focusing on metadata, ratings, and reviews for the generation and presentation of the suggestion lists. However, images describe different aspects of an item, which might be used to steer its presentation to satisfy specific information needs. We propose two user interfaces for image-based information filtering that support item exploration and comparison by analyzing the scenes described by the images, or the objects recognized in them. In a user test in the home-booking domain, we found that participants preferred scene-based information filtering to object-based and traditional visualization of items and reviews in product catalogs.**

***Index Terms*—review-based recommender systems, information filtering, human-centric computing and services**

I. Introduction

Current recommender systems use item ratings [1], features [2], and textual reviews provided by past consumers [3], [4] to suggest and present items to the user. Most of these systems manage images as black-box elements, overlooking the fact that they are a rich source of information about different aspects of items and that they can be integrated with the other, traditionally explored data to improve the visualization of recommendations. In this work, we aim to exploit the images that the user inspects as information filters. Specifically, we investigate the following research questions:

- *RQ1: are images useful to support information filtering in item lists? At which granularity level?*
  We investigate the usefulness of different aspects of images to filter item data. For instance, the photos showing the bathroom of an apartment, or a detail of its shower, might be used to focus on more or less specific types of information extracted from its metadata and reviews.
- *RQ2: does image-based filtering enhance user awareness about items and decision-making?*
  We investigate whether filtering information through image selection enhances the user's awareness about items, their comparison, and the selection of the preferred ones.

Exploiting images to enhance the interaction with the user is particularly relevant to home-booking and hotel-booking catalogs, which report several photos and reviews for each accommodation, presenting a relevant amount of details that might make comparisons and selections lengthy and burdensome. Thus, we choose the home-booking domain as a test bed for our work, and we propose two user interfaces for image-based information presentation and filtering, which we compare to a standard user interface similar to the one provided by Airbnb.[1] The results of the user study, which involved 71 participants, show that users prefer information filtering based on the contexts (scenes) of the images inspected by the user; e.g., the bedroom or surroundings of a home.

Notice that, in [5], we proposed a multimodal, service-oriented justification model that presents recommendation results based on a set of high-level evaluation dimensions of experience derived from a knowledge-intensive analysis and specification of the service underlying item fruition. In the present paper, we abstract from the service model to enhance scalability. Moreover, we focus on the impact on user awareness and decision-making of different facets of images (the types of objects they show, or the scenes they represent).

Section II presents the related work. Sections III and IV describe the data and the user interfaces. Sections V and VI outline the user study we carried out and its results. Sections VII and VIII discuss the results and close the paper.

II. Related Work

Most online catalogs show the images of each item they present without organizing the carousels based on the content provided by the individual photos. For instance, Airbnb and Booking[2] mix indoor and outdoor scenes, exposing rather long lists of images (up to several dozen) to be browsed in search of specific information. Differently, we support image-based filtering, through image analysis, to enable the user to select the relevant sets of photos to be browsed.

Our work advances review-based recommender systems [3], [4], [6], which extract and present experience data from consumer feedback, by filtering and presenting item data based on the images that the user inspects. We also bring faceted exploration support [7], [8] to multimodal information filtering by using image analysis to extract the facets of items.

Images are analyzed, e.g., to extract the features of clothes in MMFashion [9] and to identify the ingredients of food recipes in [10]. Some recommender systems employ image

[1] https://airbnb.com
[2] https://www.booking.com

analysis to build user profiles [11], or to identify sets of similar items, such as clothes similar to the user's selections, or pairing in fashion recommender systems [12], [13]. Moreover, recommender systems of images suggest which ones the user might appreciate through feature analysis [14]. Other authors combine image processing for feature enhancement and recommendation techniques to personalize the ranking of cancer drugs [15]. Furthermore, in [16], feature extraction on images is applied in a hybrid recommender system to suggest favorite restaurants through Matrix Factorization. Different from these works, we analyze the *objects* displayed in the images, and the *scenes* they represent, to identify key facets for information filtering. Specifically, we use scene and object recognition [17] to recognize the indoor and outdoor photos of a home, its rooms (bedroom, bathroom, etc.), and the presence of objects like a bed or shower.

Previous explanation [18], [19] and justification models [20], [21] support the recommendations by presenting textual information about items, possibly enriched with a visualization of quantitative data [6], [13], [22], [23]. Hybrid recommender systems [24] model different perspectives of relevance, but they strictly focus on the same types of information. Our work makes a step forward by integrating images in product presentation to enhance user awareness about item aspects brought by their photos.

III. Data

For the purpose of the user study, we used an Airbnb reviews dataset of homes located in London, which we downloaded from http://insideairbnb.com/get-the-data.html in January 2021. This dataset provides various data about each home h, like its name, a link to its host's Airbnb page, and a list of offered amenities such as WiFi and parking. The dataset also contains the mean rating received by h and the reviews written by previous guests. Finally, it provides a link to a single image of h. To obtain more than one image per home, in 2022 we scraped the Airbnb website to retrieve the photos of the homes of the dataset that were still present on the website, and we excluded the other apartments.

We then pre-processed the images of such homes and their textual information to obtain a structured representation of both data types. In the following, we describe the pre-processing tasks. To accommodate a dynamic catalog that receives new reviews and new homes, this type of analysis should be periodically carried out in a batch process.

A. Pre-processing of Images

1) Scene recognition: this analysis is aimed at recognizing the context represented by the image. To perform it, we used Places365-Standard[3], a dataset consisting of 1.8 million training and 36,000 validation images from 365 scene categories (henceforth denoted as $SCENES_P$). We accomplished the recognition of scenes using `ResNet50`[4] [25].

Places365-Standard contains scenes, such as hotel-room, volcano, and embassy, that are poorly relevant to home-booking. Therefore, we projected $SCENES_P$ onto the scenes relevant to our domain. Moreover, we mapped the similar elements of $SCENES_P$ (e.g., bedroom and child's room) to a single value. The resulting set of scenes, denoted as $SCENES$, consists of the following categories:

$SCENES$ = {kitchen, living_room, bedroom, bathroom, services, surroundings, attic, basement, studio, corridor, dining_room, dressing_room, stairs, house, patio, laundry}.

To classify the images of the homes in $SCENES$, for each image i, we worked as follows:

1) We applied `ResNet50` and retrieved a list $S = [s_1, \ldots, s_5]$, with $s_j \in SCENES_P$, sorted by likelihood score in descending order. This list represents the possible scenes described by i that the algorithm has recognized, with their degrees of certainty.
2) Then, we used the mappings to convert the elements of S into a new list $S' = [s_1, \ldots, s_k]$ ($k \leq 5$) with $s_j \in SCENES$ to obtain the classification of i in the scenes used in our experiments. If $S' \neq \emptyset$, we tagged i with $s_1 \in S'$, i.e., the most probable scene. Otherwise, we filtered out the image because it would not be possible to suitably handle it when presenting the home.

It is worth noting that S' is empty when none of the elements of S is mapped to the scenes of $SCENES$. Assuming that the images of the Airbnb homes are relevant to them, this result can be explained by a failure in scene recognition. For instance, we analyzed a small sample of failures, and we found a very elegant building that was tagged as an embassy ($\notin SCENES$). To support the automated management of images, we opted for filtering out the problematic ones. We denote as I the set of successfully classified images.

2) Object recognition: the analysis consists of extracting the types of objects displayed in the images $i \in I$. We applied object recognition to identify the entities appearing in them, e.g., a bed or a TV. To perform object recognition, we used `UODDM` (Unified Object Detection with Deep Models) [26][5], which has achieved state-of-the-art results on indoor scene understanding on SUN RGB-D [27] (a dataset for indoor scenes that contains annotations of object instances). The result of this analysis is a list of recognized classes of objects for each image $i \in I$, which we used to annotate it.

After the pre-processing, each image i used in our experiments was enriched with the following vector representation:

$$\vec{v_i} = [scene, [class_1, \ldots, class_n]]$$

where $scene$ represents the scene recognized with maximum certainty and $class_k$ represents a type of object identified in i. For example, the image (ι) of the bedroom in Figure 1 has the following vector representation:

$\vec{v_\iota}$ = ["bedroom", ["bed", "pillow", "window", "chair", "desk", "cabinet", "dresser", "picture"]]

[3]https://paperswithcode.com/dataset/places365

[4]https://github.com/CSAILVision/places365

[5]https://github.com/liketheflower/UODDM

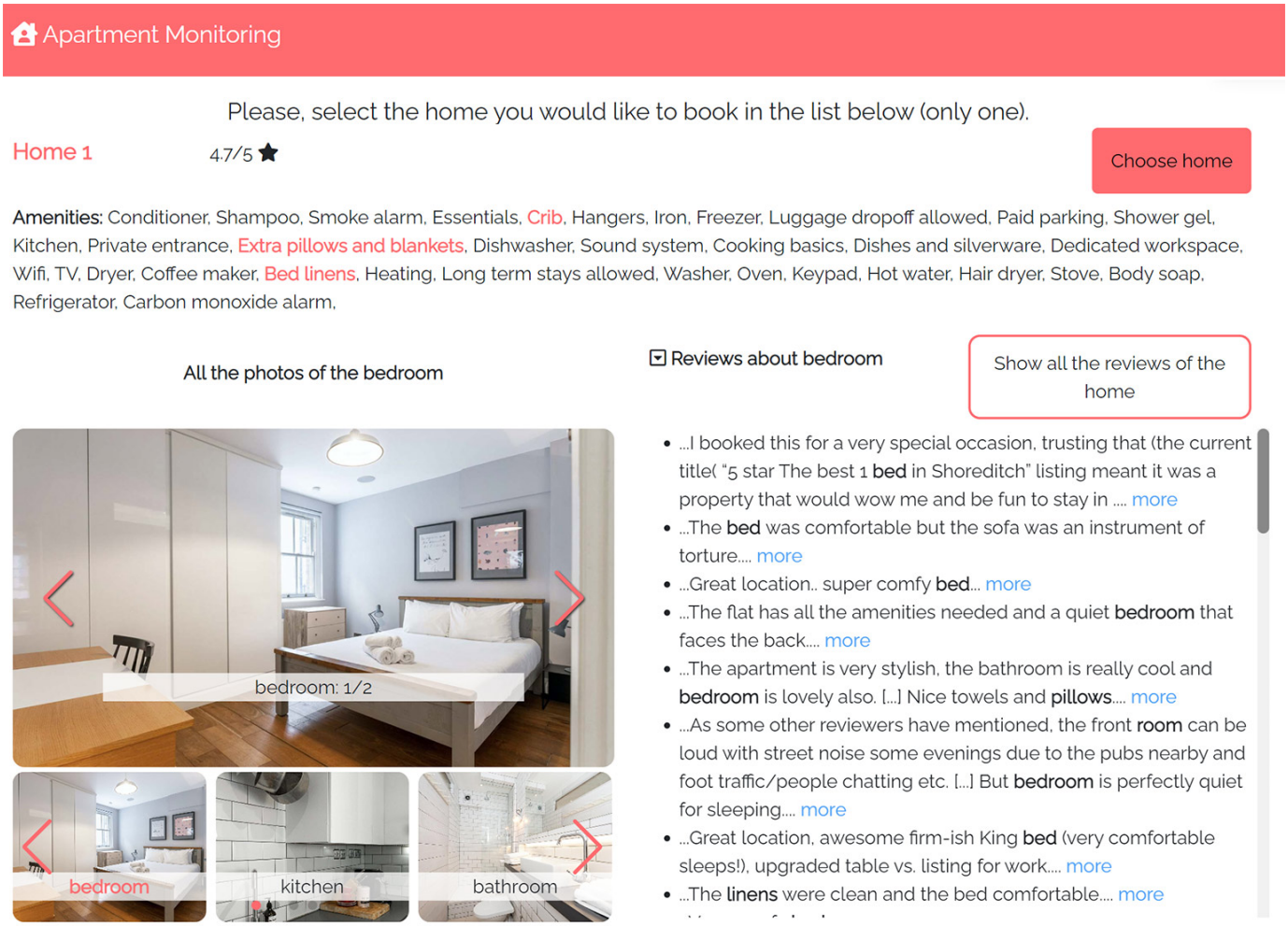

Fig. 1. FILTER-BY-SCENE user interface.

TABLE I
MAPPINGS BETWEEN SCENES AND LEMMAS, USED TO CLASSIFY AMENITIES AND REVIEWS BY SCENE (*SCENES-LEMMAS*)

Scene	Lemmas
kitchen	cooker, dishwasher, fridge, kettle, microwave, plateoven, ...
bedroom	bed, blanket, bunkbed, pillow, sheet, slipper, wardrobe, ...
bathroom	bathtub, towel, bidet, hair-dryer, shower, toiletry, ...
...	...

B. Pre-processing of Textual Data

As far as the textual information about the homes is concerned (amenities and reviews), we worked as follows:

- We lemmatized the amenities of the homes using the `spaCy` library [28].
- We filtered the English reviews using the `langdetect` library [29] to work on English reviews (R).
- For each review $r \in R$, we used a standard NLP approach, which involved lemmatization and dependency parsing, to extract its sentences and, for each sentence, the lemmas of the nouns occurring in it. For these tasks, we used `spaCy`.
- We classified the lemmas of the amenities and sentences in the elements of $SCENES$ (see Section III-A) using `spaCy` to match synonyms. This resulted in the set of *SCENES-LEMMAS* mappings; see Table I.
- We mapped the same lemmas to the classes of objects defined in $CLASSES$ to match synonyms. This resulted in the *CLASSES-LEMMAS* mappings (not shown).
- Finally, we indexed the reviews, and their individual sentences, by lemmas, and by scene, to support their retrieval during the interaction with the user.

IV. USER INTERFACES

For the purpose of the user study, we developed a web-based test application that manages the user interfaces we evaluate. In the following, we describe them and their underlying techniques. For each user interface, the application presents three homes to choose from. In the figures, we only show the first one for brevity.

A. Information Filtering by Scene

FILTER-BY-SCENE is centered around the scenes of the presented home (h); see Figure 1.

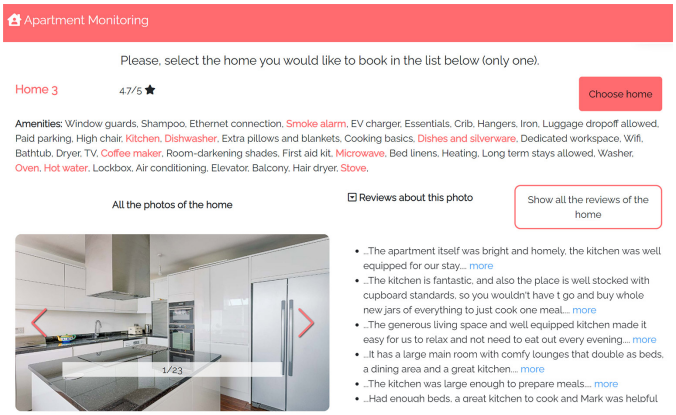

Fig. 2. FILTER-BY-OBJ user interface.

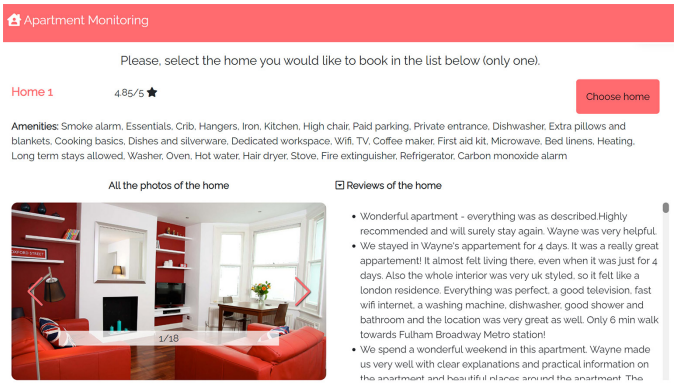

Fig. 3. BASELINE user interface.

- The main component is the widget showing the carousels of h's images. Each carousel corresponds to a scene $sc \in SCENES$ such that h has at least one photo classified in it. Thus, the number of carousels depends on how many scenes are captured in h's photos. The user can browse the list of carousels and click on a specific one to view the images of the corresponding scene.
- At the right of the carousels, the system shows the reviews of h filtered by scene to focus on the information related to the visualized images. Data is filtered through the indexing of reviews by scene; see Section III-A.
  The reviews can deal with different aspects of the home and of the stay. For instance:

  > "Apartment was great. Overall perfect location in the center of everything, easily walkable to all the sites. Beds were comfortable, showers were great. Full kitchen, but we didn't use. Washer included, which was a nice feature since I was on a multi week work trip in Europe. [...]"

  Thus, for each review r, the system only shows the sentences that are indexed in the current scene sc (carousel). The user can read the complete review r by clicking on its "more info" link. In that case, the sentences indexed by sc are highlighted in boldface. The user can also visualize all the reviews of h, or go back to those referring to sc, by clicking on the "Show all the reviews of the home" and "Reviews about ..." buttons respectively.
- Above the images, the user interface shows a fictitious name of the home[6], the mean rating it received from previous guests, and the list of the amenities it offers. The system highlights the amenities whose lemmas are classified in the current scene; see Table I.

B. Information Filtering by Objects

Figure 2 shows FILTER-BY-OBJ, which has most elements in common with FILTER-BY-SCENE but differs in the following features:

1) FILTER-BY-OBJ shows a single carousel including all the images of the home h.
2) Given the image i visualized in the user interface and its vector $\vec{v_i} = [sc, [c_1, \ldots, c_m]]$, FILTER-BY-OBJ filters the reviews of h by scene sc and promotes those that mention the types of objects c_j recognized in i.
   The selection of reviews by scene is the same as the one applied in FILTER-BY-SCENE. The identification of the reviews by objects is carried out as follows:
   - The system retrieves the object classes recognized in i from $\vec{v_i}$ and uses the *CLASSES-LEMMAS* mapping (see Section III-B) to identify the lemmas L that are relevant to i.
   - Then, the system selects the reviews of h indexed by a lemma $l \in L$. Given such reviews, it shows (and highlights when the user expands them) all the sentences indexed by a lemma $l \in L$.

C. Baseline

The BASELINE user interface, shown in Figure 3, is inspired by the Airbnb website. It presents the list of photos and reviews of the home.

V. User Study

We compared the user experience and the interaction behavior with the user interfaces presented in Section IV in a user study.[7] Our test application guided the execution of the test without our supervision. During the interaction with the users, the application logged their behavior, using numerical IDs to anonymously identify their actions. We used attention checks in the questionnaires it administered to filter out the experiments done carelessly.

We designed the user study as a within-subjects one, managing the three treatment conditions (FILTER-BY-SCENE, FILTER-BY-OBJ, and BASELINE) as independent variables. Every participant received the three treatments, in counterbalanced order to reduce the effect of fatigue and practice, and the result biases. The application did not impose any time limits

[6]We hide the real names of the homes to prevent the user from finding them on the Airbnb website.

[7]Our experiment has been approved by the Ethics Committee of the University of Torino (Protocol Number: 0421424).

TABLE II
QUESTIONNAIRE INVESTIGATING USERS' TRUST IN BOOKING SYSTEMS AND INTEREST IN REVIEWS AND IMAGES - MEAN VALUES ARE IN [1, 5]

Statement	Mean v.
S1: When comparing some homes, photos are important.	4.718
S2: When comparing some homes, reviews are important.	4.465
S3: I tend to trust the suggestions generated by booking systems.	3.126
S4: I think that the ratings given by other users are enough to book homes.	3.014
S5: I need to inspect the reviews given by other users to book homes.	4.309
S6: I need to inspect the description of the home to book it.	4.451
S7: I need to inspect the photos of the home to book it.	4.662

on the execution of the experiment and guided the participants in the following phases of interaction:

1) Description of the experiment and signing of the informed consent (https://bit.ly/42URUwE). In this phase, the application also asked the user to declare to be 18 years old or over (mandatory to continue the test).
2) Collection of demographic data, information about the user's cultural background, and familiarity with booking and e-commerce platforms.
3) Administration of the Need for Cognition questionnaire [30] to investigate the user's tendency to engage in and enjoy thinking.
4) Administration of the questionnaire in Table II.
5) Interaction with the three user interfaces, in counterbalanced order. For each one, the system presented three homes, asked the user to select the preferred one, and administered the post-task questionnaire of Table III. The statements of this questionnaire, in a 5-Point Likert scale, are taken from [31]–[33] and measure the experience and evaluation of the user interface. The questionnaire also had a free-text area for comments.
6) Administration of a very short post-test questionnaire to ask participants whether filtering data by scene, or by images, was useful.
7) Administration of the TIPI [34] questionnaire to investigate the user's personality traits.

We recruited participants through social networks, public mailing lists, and the students' mailing lists of the Computer Science courses at the University of Torino (we linked the test application in the invitation message). All the participants joined the experiment voluntarily, without any compensation.

VI. RESULTS

We carried out the user test from May 2nd to May 20, 2023; 79 people performed the test but we excluded 8 of them because they did not pass the attention checks. The mean duration of the experiment, considering the 71 participants who successfully completed the test, was about 22 minutes. According to power analysis, this sample size supports statistically significant results with α = 0.05, *power* = 0.80, and *effect size* = 0.35.

A. Participants' Data, Backgrounds and Opinions

Descriptive statistics of the 71 participants:

- *Gender*. Female (31), male (39), prefers not to answer (1).
- *Age*. 18-31 (60), 31-50 (5), > 50 (6).
- *Educational level*. High school (20), university (47), doctorate (4).
- *Background*. Scientific (33), technical (14), humanities (9), language studies (4), economics (3), other (8).
- *Familiarity with computers*. Beginner (5), average (39), advanced (27).
- *Frequency of usage of booking and e-commerce sites*. Never (2), a few times overall (7), a few times a year (27), a few times a month (35).

The participants' answers to the questionnaire of Table II show that they consider home' photos (S1, $M = 4.718$) and reviews (S2, $M = 4.465$) as very important and they moderately trust the suggestions generated by booking systems (S3, $M = 3.126$). The users declared that ratings are somehow sufficient to book homes (S4, $M = 3.014$) and they need to inspect homes' reviews (S5, $M = 4.309$), descriptions (S6, $M = 4.451$) and photos (S7, $M = 4.662$) to book apartments.

B. Post-task Questionnaire on the Whole Participants Group.

Table III shows the user experience result of the post-task questionnaire for each user interface. FILTER-BY-SCENE achieves the best results in all the statements, and most of them have a good statistical significance; the second best is FILTER-BY-OBJ.

The results with the highest statistical significance show that FILTER-BY-SCENE helped users compare homes (Q2, $p = 0.02$) and made it possible to quickly find information about them (Q8, $p = 0.002$). Moreover, FILTER-BY-SCENE helped users to understand why homes were good or bad (Q1, $p = 0.011$). People felt confident using this system to compare homes (Q12, $p = 0.035$). The other statistically significant results show that FILTER-BY-SCENE was informative (Q3) and provided enough data to make a selection (Q5), without being too much cluttered or confusing (Q4); the information about the homes was easy to interpret, and understand (Q7). Furthermore, the users declared that they would like to frequently use FILTER-BY-SCENE to compare homes (Q9). They felt more confident while using this system than the other ones.

Even though FILTER-BY-SCENE obtained mean values that are positioned in the middle-high portion of the evaluation scale, the user experience results show that it improved decision-making more than BASELINE that resembles traditional home-booking platforms. Indeed, BASELINE and FILTER-BY-OBJ expose users to a large amount of information, making item comparison difficult.

Few participants provided free-text comments in the post-task questionnaire. In the following, we summarize the most interesting ones:

- FILTER-BY-SCENE received 12 textual comments. Some people liked the fact that the images and reviews were grouped and contextualized to specific rooms of the homes. The labels concerning the scene helped these users find the information they were looking for. Overall, they appreciated this information filter.

TABLE III

POST-TASK QUESTIONNAIRE RESULTS ON THE WHOLE GROUP OF PARTICIPANTS. THE BEST VALUES FOR EACH STATEMENT ARE IN BOLDFACE (MINIMUM VALUE FOR Q4, Q6, AND Q10, MAXIMUM FOR THE OTHER STATEMENTS - MEAN VALUES ARE IN [1, 5]). WE REPORT THE P-VALUES OF THE STATISTICALLY SIGNIFICANT DIFFERENCES ACCORDING TO A KRUSKAL-WALLIS TEST

Statement	**p-value**	**FILTER-BY-SCENE**	**FILTER-BY-OBJ**	**BASELINE**
Q1: It was easy to understand why some homes were good and others not.	0.011	**3.507**	3.085	3.056
Q2: The system helped me to compare the homes.	0.002	**3.577**	3.352	3.028
Q3: The system was sufficiently informative.	0.084	**3.831**	3.761	3.535
Q4: The system was cluttered or confusing.	0.019	**2.521**	2.901	3.014
Q5: The information about the homes was sufficient for me to select a home.	0.006	**3.986**	3.761	3.563
Q6: The system provided too much information about the homes.		**2.901**	3.099	3.211
Q7: The information about the homes was easy to interpret and understand.	0.090	**3.746**	3.451	3.394
Q8: I found the information about homes quickly.	0.002	**3.831**	3.493	3.239
Q9: I think that I would like to frequently use this system to compare homes.	0.047	**3.338**	3.070	2.930
Q10: I found this system to compare homes unnecessarily complex.		**2.592**	2.859	2.901
Q11: I thought this system to compare homes was easy to use.		**3.521**	3.465	3.394
Q12: I felt very confident using this system to compare homes.	0.035	**3.549**	3.282	3.169

TABLE IV

POST-TASK QUESTIONNAIRE RESULTS GROUPED BY NEED FOR COGNITION - WE USE THE SAME NOTATION AS IN TABLE III

Statement	**Low NfC group** (NfC < 3.5; 35 participants)				**High NfC group** (NfC ≥ 3.5; 36 participants)			
	p-value	**FILTER-BY-SCENE**	**FILTER-BY-OBJ**	**BASELINE**	**p-value**	**FILTER-BY-SCENE**	**FILTER-BY-OBJ**	**BASELINE**
Q1: It was easy to understand why some homes were good and others not.		**3.543**	3.286	3.229	0.031	**3.472**	2.889	2.889
Q2: The system helped me to compare the homes.	0.043	**3.771**	3.571	3.286	0.014	**3.389**	3.139	2.778
Q3: The system was sufficiently informative.		**3.914**	3.829	3.629		**3.750**	3.694	3.444
Q4: The system was cluttered or confusing.	0.033	**2.457**	2.771	3.114		**2.583**	3.028	2.917
Q5: The information about the homes was sufficient for me to select a home.	0.026	**4.086**	3.800	3.657		**3.889**	3.722	3.472
Q6: The system provided too much information about the homes.	0.089	**2.800**	3.229	3.314		3.000	**2.972**	3.111
Q7: The information about the homes was easy to interpret and understand.		**3.829**	3.543	3.486		**3.667**	3.361	3.306
Q8: I found the information about homes quickly.	0.041	**3.857**	3.743	3.286	0.030	**3.806**	3.250	3.194
Q9: I think that I would like to frequently use this system to compare homes.		**3.486**	3.257	3.200	0.091	**3.194**	2.889	2.667
Q10: I found this system to compare homes unnecessarily complex.		**2.543**	2.686	2.914		**2.639**	3.028	2.889
Q11: I thought this system to compare homes was easy to use.		**3.629**	3.686	3.486		**3.417**	3.250	3.306
Q12: I felt very confident using this system to compare homes.		**3.914**	3.571	3.543		**3.194**	3.000	2.806

- FILTER-BY-OBJ received 11 textual comments, 8 of which highlighted that, on the one hand, the filter was too specific to quickly obtain an overview of previous guests' opinions about the homes. Some of these users criticized the fact that reviews overlapped between photos as this increased the number of textual comments to read.
- BASELINE received 11 textual comments. Most participants complained that it provided too many reviews, representing unstructured information, without any filtering support. Moreover, people complained that it's difficult to overview and compare homes.

To investigate the impact of participants' personalities on the experience with the three user interfaces, we analyzed the results of the post-task questionnaire by splitting the sample of users by Need for Cognition (NfC), and by personality traits (TIPI questionnaire).

C. Post-task Questionnaire - Split by Need for Cognition

Table IV shows the user experience results obtained by dividing the sample of participants by their Need for Cognition. The low-NfC group includes 35 people having NfC < 3.5. The other group includes 36 people having NfC ≥ 3.5. The goal of this analysis is to understand if the tendency to engage in and enjoy thinking impacts the perception of the visual information filters we propose.

The results are consistent with the findings concerning the whole participants' sample, with a lower number of statistically significant differences. Both subgroups appreciated FILTER-BY-SCENE more than the other user interfaces. However, the people having low NfC evaluated FILTER-BY-SCENE higher than the overall group of participants. Conversely, the other subgroup gave slightly lower scores.

D. Post-task Questionnaire - Split by Personality Traits

To investigate the impact of personality traits on user experience, we divided the sample of users into subgroups using 4 as a threshold for the splits (≤ 4, and > 4).

- First, we consider the *openness* trait, which relates to being inventive, curious, and willing to try new experiences. The participants having a high score in this trait preferred FILTER-BY-SCENE to FILTER-BY-OBJ in all the aspects addressed by the post-test questionnaire (with

TABLE V
LOG ANALYSIS: "TIME" IS THE MEAN NUMBER OF MINUTES USERS SPENT ON A USER INTERFACE; "#REVIEWS" ("#IMAGES") IS THE MEAN NUMBER OF VISUALIZED REVIEWS (CLICKED IMAGES)

Event	FILTER-BY-SCENE	FILTER-BY-OBJ	BASELINE
Time (minutes)	3.1	2.612	1.771
#Reviews (visualized)	94.254	79.746	21.746
#Images clicked	32.944	32.803	26.62

5 statistically significant differences). Moreover, users preferred FILTER-BY-OBJ to BASELINE in the statements related to item comparison and exploration (Q1, Q2, Q5, Q8, $p < 0.03$). Differently, they considered FILTER-BY-OBJ as more cluttered than BASELINE (Q4, $p = 0.026$). Concerning the participants with low openness, the situation is somehow mixed, with some preference for FILTER-BY-OBJ. However, FILTER-BY-SCENE is the best-performing user interface in Q2 and Q7, which are the only two statistically significant results ($p < 0.1$).

- We then consider participants' *conscientiousness*, which refers to their desire to be diligent, careful, and self-disciplined, and is related to planning. These results might provide insights into the interest in specific data about homes for decision-making. Similarly to *agreeableness*, both subgroups of participants preferred FILTER-BY-SCENE in all the statements of the questionnaire. Specifically, the highly conscientious participants perceived that the information about the homes provided by FILTER-BY-SCENE was definitely sufficient to select a home (Q5, 4.019, $p = 0.042$). On the statistically significant results, FILTER-BY-OBJ outperforms BASELINE.
- Concerning *extroversion*, *agreeableness*, and *neuroticism* both subgroups of participants preferred FILTER-BY-SCENE. Moreover, FILTER-BY-OBJ is the second best on the statistically significant results.

E. Post-test Questionnaire

Participants evaluated the filtering of reviews by scenes as more useful ($M = 4.253$) than the filtering by objects ($M = 3.648$) but we can say that they appreciated both of them.

F. Log Analysis

As shown in Table V, the log analysis reveals that participants spent more time interacting with FILTER-BY-SCENE than with FILTER-BY-OBJ. Moreover, they spent the least time on BASELINE. This is consistent with the amount of information they explored: they visualized about 94 reviews when using FILTER-BY-SCENE, 80 with FILTER-BY-OBJ, and 22 with BASELINE. Differently, they visualized about the same number of photos (about 10 images per home in FILTER-BY-SCENE and FILTER-BY-OBJ, 9 in BASELINE).[8] These findings suggest that, regardless of the user interface they interacted with, people privileged homes' images in the comparison and selection tasks. However, they explored much more textual information when supported by the scene-based filter (and a bit less with the object-based one) than in the baseline condition.

[8]To estimate the visualization of reviews and images, we tracked the time that each of them was displayed on the screen and considered only the visualizations lasting more than 2 seconds.

VII. DISCUSSION OF RESULTS

The results of the user study show that participants preferred FILTER-BY-SCENE to FILTER-BY-OBJ and BASELINE. Moreover, the people having low NfC particularly liked it, probably because it strongly supports information filtering. It also emerged that FILTER-BY-OBJ has been perceived as better than BASELINE in several evaluation perspectives. Based on these findings, we answer our research questions as follows:

- Regarding *RQ1*, item images are very useful for focusing the information about items but the granularity level of the filter matters. FILTER-BY-SCENE, which supports coarser-grained filtering by scene, was the most appreciated user interface. Differently, the objects recognized in the images seem to be less effective.
- As far as *RQ2* is concerned, the user experience results confirm that scene-based filtering supports both user awareness about items and confidence in the selection decisions more strongly than filtering by the objects recognized in the images.

Thus, we conclude that FILTER-BY-SCENE could represent a new way to show multimodal information about homes by offering a compact visual representation to group photos by context (room, services, etc.) and selecting the textual information and feedback about the homes to view.

A limitation of the present work is the relatively low evaluation of the three user interfaces that, combined with some comments provided by participants in the free-text questions, suggest simplifying them to enhance their usability. In our future work, we plan to improve this aspect of FILTER-BY-SCENE, which we choose to further explore multimodal information filtering. We also plan to test our revised system with a larger number of participants and in other application domains, such as hotel booking, and the recommendation of restaurants, to test the applicability of our approach.

VIII. CONCLUSIONS

We proposed two user interfaces that support image-based information filtering to focus the presentation of items (homes) on specific perspectives. FILTER-BY-SCENE filters the textual information by scene (e.g. bedroom, kitchen, etc.), and FILTER-BY-OBJ applies the filter according to the objects recognized in the images that the user inspects. We compared these interfaces with a baseline that resembles traditional home-booking platforms. In a user study involving 71 participants, we found that people prefer scene-based information filtering to explore and compare homes, encouraging its use to support information exploration.

ACKNOWLEDGMENT

This work has been funded by the University of Torino.

REFERENCES

[1] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '00. New York, NY, USA: Association for Computing Machinery, 2000, p. 241–250. [Online]. Available: https://doi.org/10.1145/358916.358995

[2] Y. Deldjoo, M. Ferrari Dacrema, M. Constantin, H. Eghbal-zadeh, S. Cereda, M. Schedl, B. Ionescu, and P. Cremonesi, "Movie genome: alleviating new item cold start in movie recommendation," *User Modeling and User-Adapted Interaction*, vol. 29, pp. 291–343, 2019.

[3] M. Hernández-Rubio, I. Cantador, and A. Bellogín, "A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews," *User Modeling and User-Adapted Interaction*, vol. 29, no. 2, pp. 381–441, 2019. [Online]. Available: https://doi.org/10.1007/s11257-018-9214-9

[4] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: the state of the art," *User Modeling and User-Adapted Interaction*, vol. 25, no. 2, pp. 99–154, 2015. [Online]. Available: https://doi.org/10.1007/s11257-015-9155-5

[5] Z. F. Hu, N. Mauro, G. Petrone, and L. Ardissono, "Service-based presentation of multimodal information for the justification of recommender systems results," in *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 46–53. [Online]. Available: https://doi.org/10.1145/3565472.3592962

[6] L. Chen and F. Wang, "Explaining recommendations based on feature sentiments in product reviews," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ser. IUI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 17–28. [Online]. Available: https://doi.org/10.1145/3025171.3025173

[7] M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Yee, "Finding the flow in web site search," *Communications of the ACM*, vol. 45, no. 9, pp. 42–49, Sep. 2002. [Online]. Available: http://doi.acm.org/10.1145/567498.567525

[8] N. Mauro, L. Ardissono, and M. Lucenteforte, "Faceted search of heterogeneous geographic information for dynamic map projection," *Information Processing & Management*, vol. 57, no. 4, p. 102257, 2020. [Online]. Available: https://doi.org/10.1016/j.ipm.2020.102257

[9] X. Liu, J. Li, J. Wang, and Z. Liu, "MMFashion: An open-source toolbox for visual fashion analysis," in *Proceedings of the 29th ACM International Conference on Multimedia*, ser. MM '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 3755–3758. [Online]. Available: https://doi.org/10.1145/3474085.3478327

[10] Y. Kawano, T. Sato, T. Maruyama, and K. Yanai, "[demo paper] mirurecipe: A mobile cooking recipe recommendation system with food ingredient recognition," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–2.

[11] R. Kitamura and T. Itoh, "Tourist spot recommendation applying generic object recognition with travel photos," in *2018 22nd International Conference Information Visualisation (IV)*, 2018, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8564129

[12] S. Chakraborty, M. S. Hoque, N. Rahman Jeem, M. C. Biswas, D. Bardhan, and E. Lobaton, "Fashion recommendation systems, models and methods: A review," *Informatics*, vol. 8, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2227-9709/8/3/49

[13] Y. Deldjoo, F. Nazary, A. Ramisa, J. Mcauley, G. Pellegrini, A. Bellogin, and T. Di Noia, "A review of modern fashion recommender systems," 2022. [Online]. Available: https://arxiv.org/abs/2202.02757

[14] K. Kobyshev, N. Voinov, and I. Nikiforov, "Hybrid image recommendation algorithm combining content and collaborative filtering approaches," *Procedia Computer Science*, vol. 193, pp. 200–209, 2021, 10th International Young Scientists Conference in Computational Science, YSC2021, 28 June – 2 July, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921020615

[15] L. Brandão, F. Belfo, and A. Silva, "Wavelet-based cancer drug recommender system," *Procedia Computer Science*, vol. 181, pp. 487–494, 2021, CENTERIS/ProjMAN/HCist 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921002362

[16] W.-T. Chu and Y.-L. Tsai, "A hybrid recommendation system considering visual information for predicting favorite restaurants," in *Proceedings of the 26th Int. Conf. on World Wide Web*, ser. WWW '17. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017, p. pages 1313–1331.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jun 2016, pp. 779–788. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.91

[18] N. Tintarev and J. Masthoff, "Evaluating the effectiveness of explanations for recommender systems," *User Modeling and User-Adapted Interaction*, vol. 22, no. 4-5, pp. 399–439, 2012.

[19] F. Gedikli, D. Jannach, and M. Ge, "How should I explain? a comparison of different explanation types for recommender systems," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 367–382, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581913002024

[20] N. Mauro, H. Zhongli Filippo, and L. Ardissono, "Justification of recommender systems results: a service-based approach," *User Modeling and User-Adapted Interaction*, vol. 33, no. 3, pp. 643–685, 2022. [Online]. Available: https://doi.org/10.1007/s11257-022-09345-8

[21] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 188–197. [Online]. Available: https://www.aclweb.org/anthology/D19-1018

[22] M. Millecamp, N. N. Htun, C. Conati, and K. Verbert, "What's in a user? Towards personalising transparency for music recommender interfaces," in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 173–182. [Online]. Available: https://doi.org/10.1145/3340631.3394844

[23] A. El Majjodi, A. D. Starke, and C. Trattner, "Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system," in *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 48–56. [Online]. Available: https://doi.org/10.1145/3503252.3531312

[24] B. Cardoso, G. Sedrakyan, F. Gutiérrez, D. Parra, P. Brusilovsky, and K. Verbert, "IntersectionExplorer, a multi-perspective approach for exploring recommendations," *International Journal of Human-Computer Studies*, vol. 121, pp. 73 – 92, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581918301903

[25] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[26] X. Shen and I. Stamos, "Unified Object Detector for different modalities based on vision transformers," 2023.

[27] S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[28] Explosion AI, "spaCy - industrial Natural Language Processing in python," 2017, https://spacy.io/.

[29] N. Shuyo, "langdetect," 2020, https://pypi.org/project/langdetect/.

[30] G. Coelho, P. H. P. Hanel, and L. J. Wolf, "The very efficient assessment of need for cognition: Developing a six-item version," *Assessment*, vol. 27, no. 8, pp. 1870–1885, 2020. [Online]. Available: https://doi.org/10.1177/1073191118793208

[31] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, ser. RecSys '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 157–164. [Online]. Available: https://doi.org/10.1145/2043932.2043962

[32] C. Di Sciascio, P. Brusilovsky, C. Trattner, and E. Veas, "A roadmap to user-controllable social exploratory search," *ACM Transaction on Interactive Intelligent Systems*, vol. 10, no. 1, aug 2019. [Online]. Available: https://doi.org/10.1145/3241382

[33] J. R. Lewis and J. Sauro, "The factor structure of the System Usability Scale," in *Human Centered Design*, M. Kurosu, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 94–103.

[34] S. Gosling, P. Rentfrow, and W. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092656603000461