User Intentions in Digital Photo Production: A Test Data Set

Mathias Lux, Desara Xhura, and Alexander Kopper

Klagenfurt University, Klagenfurt Austria
{mlux,dxhura,akopper}@itec.uni-klu.ac.at

Abstract. Taking a photo with a digital camera or camera phone is a process triggered by a certain motivation. People want for instance to document the progress of a task, others want to preserve a moment of joy. In this contribution we present an openly available dataset with 1,309 photos along with annotations specifying the intentions of the photographers. This data set is the result of a large survey on Flickr and shall provide a common basis for joint research on user intentions in photo production. The survey data was validated using Amazon Mechanical Turk. Besides discussing the process of creating the data set we also present information of the structure and give statistics on the data set.

Keywords: User Intentions, Digital Photos.

1 Introduction

Researchers in multimedia information systems, visual information retrieval and information retrieval in general have lately put more and more emphasis on research regarding users' context. A common definition of context is: *Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.* [1]. A user's intention – defined as *a thing intended; an aim or plan*[1] – therefore is part of a user's context.

In multimedia information systems user intentions are manifold. In search scenarios users might want to find multimedia data to gain knowledge, or to entertain themselves. In publishing scenarios users might intend to communicate ideas or share feelings with others. To learn about intentions, the users have to answer *why* they want to search, share, or store a video or image. Figure 1 shows two images. Image (A) has been taken to preserve a bad feeling. The photographer noted: "because i [sic] was feeling sad at that time and everything seems as sharp and hard to me as the endins [sic] of this plant.". Image (B) on the other hand was taken for a functional reason. The photographer claimed: "I'm origami folder, and I took this photo to archive my work and share it with other origami folders."

[1] Oxford Dictionaries, `http://oxforddictionaries.com/`

C. Gurrin et al. (Eds.): MMM 2014, Part I, LNCS 8325, pp. 172–182, 2014.

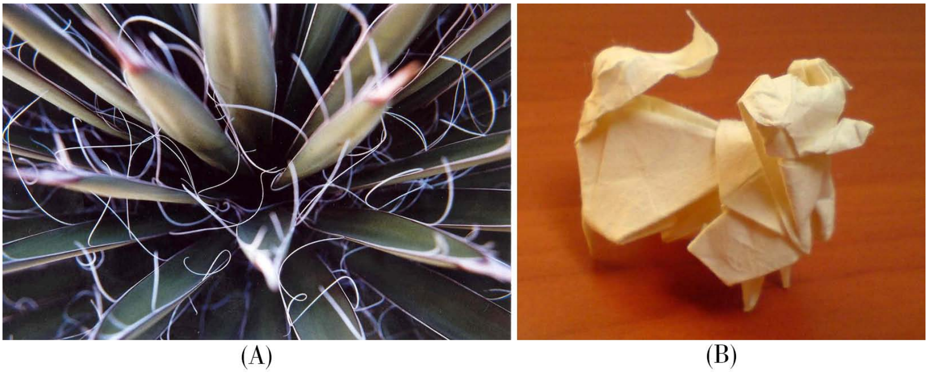

Fig. 1. Two sample photos from our test data set taken with different intentions

In this sense we have created a data set, where 1,309 images, shared on the internet, have been annotated by their owners to indicate why these images have been taken. The images were randomly selected from the Flickr web site and their publishers have been contacted to take part in a survey. An additional, crowd-sourced verification step was done with the help of Amazon Mechanical Turk. The data set is publicly available for scientific use under Creative Commons Attribution License[2]. Note at this point that this is not a test data set in the common sense of multimedia retrieval. There are neither queries and topics given for the data set, nor can it considered being a ground truth. Its value is (i) its nature of being a first data source for research on user intentions in multimedia, and (ii) that the data set provides a basis common to different research groups due to its open nature. Its nature is comparable to the infamous *AOL search log data*, where also no topics were given, but the data set was appreciated (in terms of availability of data not in terms of releasing it without asking the users) by the research community. However, in our case we asked the photographers for permission to release the data.

This paper describes the data set starting with a short overview on related work and research on user intentions. Then the acquisition process is outlined and basic statistics and information on the data set are given. We conclude the paper with a discussion on the impact of the data set and give an outlook on future work.

2 Related Work

Data sets for multimedia retrieval and computer vision have quite a long history, as it is commonly agreed that building on each others research results can only work if methods and data are made available. A discussion on the Corel data set, which was employed often in sub sets, and its implications are presented

[2] Note that the URL is not given to the double blind review process.

in [2]. Today, a well-known and well-received data set is for instance the MIR-Flickr [3] data set. Since 2010 it provides 1,000,000 images from Flickr along with metadata including tags, title, license and EXIF. Other examples – just to pick a few out of many – are the Caltech-256 Object Category Data set [4], which consists of more than 30,000 images in 256 categories, and the PASCAL data set, which was developed for the PASCAL Visual Object Classes (VOC) Challenge [5].

The problem of capturing the intention of multimedia information system users is diverse, so different approaches have been tried. A preliminary survey on the creation of videos has been presented in [6]. Similarly [7], [8], [9] investigate the intentions people have for capturing photos with phone cameras and [10] investigates intentions for capturing photos independently from the camera used. Intentions for watching online videos have been investigated in [11]. As a part of a survey on user (sub-)groups in multimedia information systems, the goal-directedness of users is investigated in [12]. User intentions for searching images are discussed in [13], where also a taxonomy of user intentions for image retrieval is presented. A taxonomy on intention classes for online video search is discussed in [14]. An application of the research on user intentions for image search is discussed in [15], where the result view of Flickr is adapted to the automatically detected search intention class.

3 Methodology and Acquisition

To collect the data set, we developed an online survey tool, which is able to (i) download recently uploaded images along with the associated metadata from Flickr, (ii) create a questionnaire for every image and (iii) invite the photographers to the survey to fill in their individual questionnaire.

In the survey, the photographers were asked

1. if their image and associated metadata could be used for non-commercial, scientific research,
2. to give additional tags describing the image content, and
3. to provide information about their intentions for taking the image.

To capture the photographers' intentions we asked the photographers to write a free text description on their motivation for taking the image and to rate their intention on some predefined intention classes.

To determine the predefined intention classes we analyzed the results presented in [7], [8], [9] and [10]. What all these studies have in common is that they distinguish between images captured for emotional or functional reasons, between images captured for personal use, to be shared between a known group of people and those to be made public, as well as between images produced for archiving memories or to be kept for a short time. This led to a list of five main intentions,

- preserve an emotion
- support a task
- recall a situation
- share with family and friends.
- publish online

We used the class *preserve an emotion* as a mean to evaluate if the given ratings were not randomly assigned. This was done by decomposing this class in two contradictory subclasses *preserve good feeling* and *preserve bad feeling*. The resulting six predefined intentions were presented to the photographers with the following textual descriptions:

- I took the photo to capture the moment or recall a specific situation later on.
- I took the photo to preserve a good feeling (luck, joy, happiness etc.).
- I took the photo to publish it online.
- I took the photo to show it to my family and friends.
- I took the photo to support a task of mine (archive or document work or task, communicate work progress etc.).
- I took the photo to capture a bad feeling (sadness, anger, depression etc.).

The photographers were asked to rate these intentions based on a five point Likert scale including {*strongly disagree*, *disagree*, *neutral*, *agree*, *strongly agree*}

The survey was active from June to September 2011. During this time we downloaded 17,119 images and sent 13,583 invitations to photographers. Out of those, 1,309 were completed, which results in a return rate of 9,6%. The invitations were delivered using the public Flickr API and posted as comments to the images. For this task Flickr user accounts were created. Flickr's anti spam restrictions deleted these accounts three times.

To increase the quality of the collected data and to gather more information about the images, we validated the results using the Amazon Mechanical Turk[3] (AMT) marketplace, where people, called AMT workers or *turkers*, fill out surveys or solve small tasks, called human intelligence tasks or *HITs*, for a small amount of money. For each image, a HIT was created, which was then presented to 5 different turkers. In the HIT *turkers* were shown the image. In a first step they had to remove tags that have no relation to the image content, then add additional tags that describe the picture and then rate the degree of manipulation of the image (natural to artifical image). After that, the free text description of phototgrapher's intention was displayed. The turkers had to rate the readability and if an intention can be inferred from it. Also, turkers had to consider the free text description and think about why the photographer captured the photo. Then they had to rate the same six predefined intention classes like the photographers did. The turkers did not see the intention ratings by the photographers.

[3] `http://www.mturk.com`

The creation of the HIT was an iterative process including several pretests. A first offline pretest employed a convenience sample of 5 people for the comprehensibility of the HITs' questions. Further pretests were undertaken on AMT with actual turkers.

The evaluation of the data set was conducted in February 2012 and lasted one month. 6,545 HITs (five for each photo) were successfully completed by 177 turkers. By manual quality control (i) 321 HITs were rejected and republished for other workers to complete and (ii) eight Turkers were blocked due to their bad working performance. The completion of a HIT was rewarded with 0.05$. In total, we spent about 360$ for the validation including the 10% fee issued by AMT. Expenses for additional pretests are not included in this calculation.

To maintain quality in the results of turkers, (i) a large portion of the HITs were reviewed manually, (ii) each HIT was completed by five different turkers and (iii) turkers could only work on HITs if their approval rate of turkers was 95% and above and they had at least 100 HITs approved in their history as AMT workers.

4 Data Set

The resulting data set consists of 1,309 samples. Each sample contains information about the image collected from three main sources: (i) taken from Flickr's API including EXIF metadata, (ii) added from the photographer in the course of the survey and (iii) added by the turkers in the HITs. An example for the

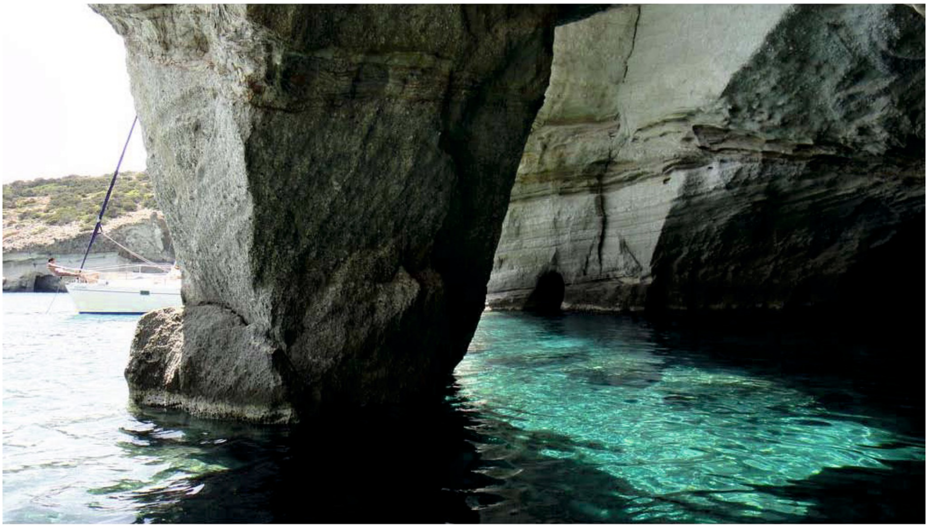

Fig. 2. Sample image from the data set. The photographer described the intention for taking the photo as "a reminder of the beautiful Island were [sic] my father came from".

Table 1. Example of a data item from the data set giving the rating on the image shown in Figure 2. A value of -2 corresponds to *strongly disagree* on the Likert scale, while a value of 2 denotes *strongly agree.*

	Photogr.	**Turkers**
Recall situation:	2	2, 0, 1, 0, 2
Preserve good feeling:	2	-2, 1, 0, 0, 1
Publish online:	2	0, 0, 0, 1, 2
Show to family & friends:	2	1, 2, 1, 1, 0
Support task of mine:	0	-2, 1, 1, 1,-2
Preserve bad feeling:	-2	0,-2, 0,-2,-2
Degree of manipulation:	-	-1,-2,-1, 0,-2
Readability:	-	-2, 2, 0, 0, 2
Infer intention:	-	0, 2, 1, 0, 0

ratings of an instance from the data set is given in Table 1. The ratings were given to the image shown in Figure 2.

Using the IP addresses of the survey participants and turkers logged by our web server, we were able to assign locations to survey participants and turkers and therefore, to get a rough idea about the originating country. The survey participants – the actual photographers – are spread over 95 different countries. Around 38% of the participants were from English speaking countries like USA, UK and Australia.

In contrast to the widespread distribution of photographers from all over the world, the majority of turkers – the people doing the validation on AMT – were from India and only a small percentage from other countries. Figure 3 gives an overview on the absolute number of participants from the six top countries (on top) and an overview on the turkers' locations (bottom). A trend of an increase of Indian turkers on AMT was already noticed by Ross et al. [16] in 2009. They observed that the share of Indian workers went from 5% in November 2008 to 36% in November 2009, so the distribution of turkers in our survey is not too unusual.

A first and pressing question was to what degree the employed intention classes were redundant. Therefore we investigated if the 6 classes were correlated in a pair wise manner. Table 2 shows the correlation matrix. Most interesting correlations are to be found between the intentions *show to family and friends*, *recall situation* and *preserve good feeling*, and that the highest correlation is between *preserve good feeling* and *recall situation* with a value of 0.45. The rest of the correlations coefficients are too small to talk of a reasonable correlation. However, the actual values indicate that with the given data sets the 6 classes of intentions are not pair wise redundant and therefore, cannot be removed.

4.1 Inter Rater Agreements

With the validation on AMT by 5 turkers for each instance the question whether the turkers agree is obvious. For quantizing inter-rater agreement we chose

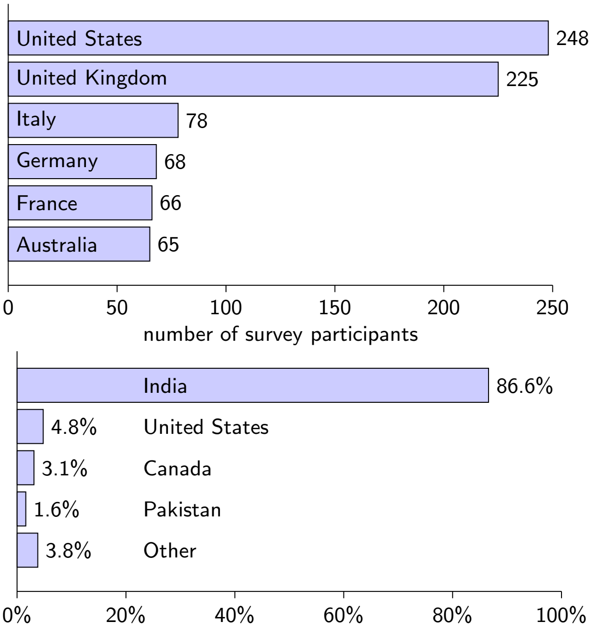

Fig. 3. Locations of the survey participants (photographers, top graph) and the turkers (workers on AMT in the validation step, bottom graph)

Table 2. Intentions Correlation Matrix

Attr.	recall	good	pub.	show	task	bad
recall	1	**0.45**	0,01	**0.29**	-0.08	-0.05
good		1	0.04	**0.29**	-0.05	-0.01
pub.			1	0.19	0.21	0.01
show				1	-0.08	-0.08
task					1	0.14
bad						1

Krippendorff's Alpha α, specifically the *R* implementation provided by the *irr* package.

Krippendorff's Alpha α is a reliability coefficient that measures the agreement between raters by taking into consideration the agreement for randomly assigned ratings. The following formula describes how the α coefficient is calculated:

$$\alpha = 1 - \frac{Observed\ disagreement}{Expected\ disagreement\ for\ random\ assignements} \tag{1}$$

Values very close to 0 indicate that the inter-rater agreement reliability is low as their ratings are very similar to random assignments and values close to 1 indicate a high agreement between raters. Negative values might occur as well, this often indicates that raters are systematically disagreeing with each other. Detailed description of how to calculate the *observed disagreement* and the *random disagreement* can be found in [17].

We used Krippendorff's Alpha α coefficient to investigate in how far the raters agreed on the rating for the six intention classes and compared that to the inter-rater agreement on (i) readability, (ii) possibility to infer intention classes based on the text description, and (iii) the degree of manipulation of the image. Figure 4 shows the histogram of Krippendorff's Alpha α for all 1,309 instances. In the left graph, showing a histogram of α for the six intention classes, α is generally lower with a large part of the values in $[0.0, 0.4]$. The agreement on the other classes is better as there are more values in $[0.4, 0.8]$. Also the plot in Figure 5 shows that the inter rater agreement follows a different distribution for the intention classes and the other ones.

Table 3. Descriptive statistics for α indicating the inter rater agreement for the six intention classes compared to the three other classes

	Intentions	**Other**
mean	0.1467	0.2321
variance	0.0316	0.0693
minimum	-0.2361	-0.2291
maximum	0.7096	0.9437

Table 3 shows statistics on α. It is easy to see that the turkers' agreement is worse for the intention classes. Of course these ratings can be considered highly subjective as the turkers had to rate the intention of the photographer, which is in many cases only partially expressed in the free text given by the photographers. Still, the inter rater agreement on the other three classes is somewhat discouraging. All in all the inter rater agreement of the turkers is low. Therefore, we assume to take a closer look at data cleaning methods before employing the data set to sort out instances that do not have the necessary information quality.

Table 4 shows statistics about the agreement of the five pretesters in the offline pretest. Compared to the Turkers' results, the participants of the pretest show a higher agreement. This intuitively sounds right as the people participating in the pretest were more likely to do a good job as they were watched and moderated while they were doing the tasks. Therefore, these values give us an

Table 4. Descriptive statistics for α from the pretest

	Intentions	**Other**
mean	0.5707	0.5104
variance	0.0157	0.1337
minimum	0.4330	-0.0499
maximum	0.7710	0.8571

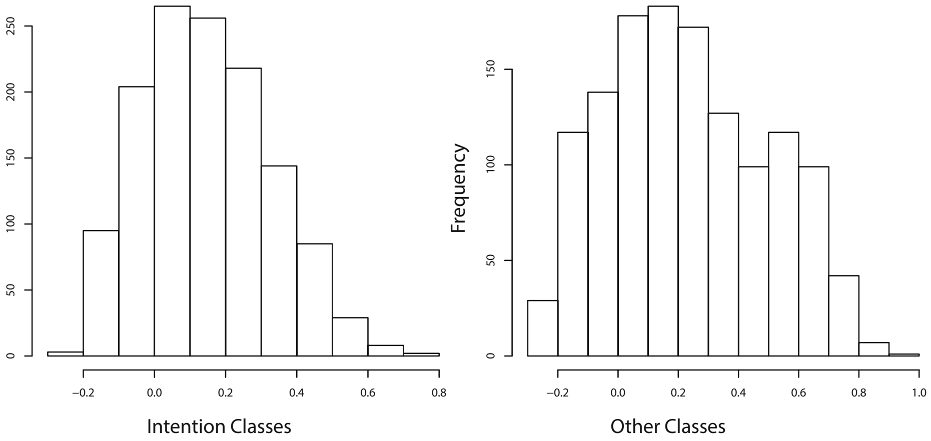

Fig. 4. Histograms of α indicating the inter rater agreement for the six intention classes compared to the three other classes

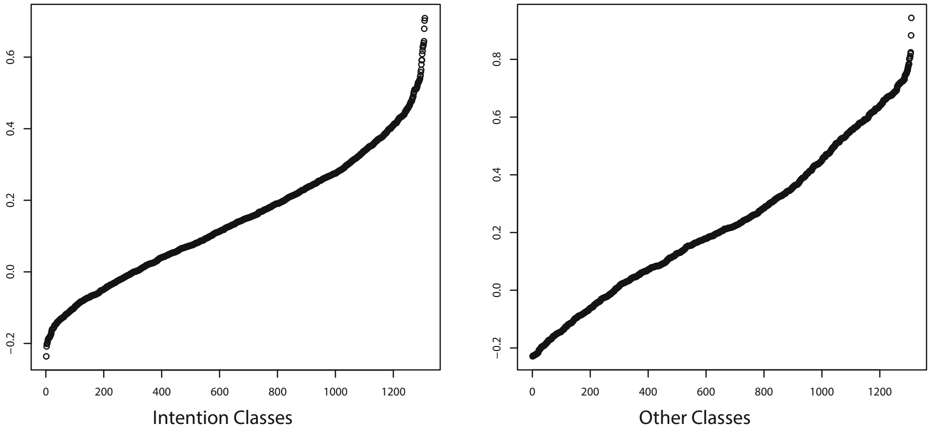

Fig. 5. Plot of the ranked values for α with the rank in the x-axis and α on y indicating the inter rater agreement for the six intention classes compared to the three other classes

idea an agreement on what issues might be possible regardless of the subjectivity in the nature of the task.

We hypothesize three main possible causes for the low inter rater agreement of the turkers.

1. **Subjective interpretation.** The turkers can have a different opinion on the the ratings.
2. **Carelessness.** Cases have been reported in literature that some turkers take short cuts and insert random values. This generates noise and lowers agreement scores.

3. **Judging on an unknown intention.** As it is already a hard task for the photographer to give an abstraction of the cause for taking the photo, a person who does not even know the photographers and their context has a hard time to judge upon intentions just knowing the photo and a few lines of text.

5 Conclusions

In this paper we have presented a data set of 1,309 photos. These photos were collected from Flickr and the photographers participated in a survey that tried to find out why the images have been taken. The data set is to this date – to the best of our knowledge – the only openly available data set[4] dealing with user intentions in multimedia. We consider this as one of the first steps towards joint research in user intentions in multimedia information systems on a common basis. While providing anecdotal evidence on actual intentions for taking photos, also text mining and pattern analysis on the data might lead to insights on why people actually take photos and put them online. Ultimately this understanding will help in providing better tools and algorithms for multimedia search, retrieval, distribution, storage and communication.

While the data set is a great tool to leverage understanding of user intentions in creating digital photos, there are several shortcomings. First of all the data set only includes photos that have already been shared and are available to the public. Hence, the data set is biased towards a sharing intention. Also the actual intention is hard to find, even for the original photographer or uploader of the image. This additional step of abstraction is something users do not appreciate. In face to face interviews we often heard the answer “I don’t know”. We assume that those, that were not willing to formulate their explicit intention for taking the photo either aborted the study or did not even start it. Still, there are multiple answers, that do not define the intention, but explain the content of the image. Furthermore the data set is rather noisy. Instances with rich information are mixed with instances that are most likely fakes or random answers.

In the near future we want to investigate the data set in full detail. First steps towards using the data set to infer photographers’ intentions have shown promising results. Also manual selection of a sub set with richly annotated instances is a next, crucial step.

References

1. Abowd, G.D., Dey, A.K.: Towards a better understanding of context and context-awareness. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 304–307. Springer, Heidelberg (1999)
2. Müller, H., Marchand-Maillet, S., Pun, T.: The truth about corel - evaluation in image retrieval. In: Lew, M., Sebe, N., Eakins, J.P. (eds.) CIVR 2002. LNCS, vol. 2383, pp. 38–49. Springer, Heidelberg (2002)

[4] Note that the URL is not given due to the double blind review process.

3. Huiskes, M.J., Lew, M.S.: The mir flickr retrieval evaluation. In: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval (MIR 2008). ACM, New York (2008)
4. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical report, California Institute of Technology (2007)
5. Everingham, M., Gool, L., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision 88, 303–338 (2010)
6. Lux, M., Huber, J.: Why did you record this video? an exploratory study on user intentions for video production. In: 2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 1–4 (May 2012)
7. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: The ubiquitous camera: An in-depth study of camera phone use. IEEE Pervasive Computing 4(2), 42–50 (2005)
8. Mäkelä, A., Giller, V., Tscheligi, M., Sefelin, R.: Joking, storytelling, artsharing, expressing affection: a field trial of how children and their social network communicate with digital images in leisure time. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 548–555. ACM (2000)
9. Van House, N., Davis, M., Ames, M., Finn, M., Viswanathan, V.: The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1853–1856. ACM (2005)
10. Lux, M., Kogler, M., del Fabro, M.: Why did you take this photo: a study on user intentions in digital photo productions. In: Proceedings of the 2010 ACM Workshop on Social, Adaptive and Personalized Multimedia Interaction and Access (SAPMIA 2010), pp. 41–44. ACM, New York (2010)
11. Lagger, C., Lux, M., Marques, O.: Which video do you want to watch now? development of a prototypical intention-based interface for video retrieval. In: Workshop on Multimedia on the Web, pp. 45–48 (2011)
12. Kemman, M., Kleppe, M., Beunders, H.: Who are the users of a video search system? classifying a heterogeneous group with a profile matrix. In: 2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 1–4 (May 2012)
13. Lux, M., Kofler, C., Marques, O.: A classification scheme for user intentions in image search. In: CHI 2010 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2010), pp. 3913–3918. ACM, New York (2010)
14. Hanjalic, A., Kofler, C., Larson, M.: Intent and its discontents: The user at the wheel of the online video search engine. In: Proceedings of the ACM 21 International Conference on Multimedia 2012, Nara, JP (November 2012)
15. Kofler, C., Lux, M.: Dynamic presentation adaptation based on user intent classification. In: Proceedings of the 17th ACM International Conference on Multimedia, MM 2009, pp. 1117–1118. ACM, New York (2009)
16. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers?: shifting demographics in mechanical turk. In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2010, pp. 2863–2872. ACM, New York (2010)
17. Krippendorff, K.: Computing krippendorff's alpha reliability. Departmental Papers (ASC) 43 (2007)