Cultural difference in image tagging

2 authors, including:

Wei Dong
Cornell University
18 PUBLICATIONS   **173** CITATIONS

Cultural Difference in Image Tagging

Wei Dong and Wai-Tat Fu

Applied Cognitive Science Lab
University of Illinois at Urbana-Champaign
405 N. Mathews Avenue, Urbana, IL 61801
wdong@illinois.edu

ABSTRACT

Do people from different cultures tag digital images differently? The current study compared the content of tags for digital images created by two cultural groups: European Americans and Chinese. In line with previous findings on cultural differences in attentional patterns, we found similar cultural differences in the order of the image parts (e.g., foreground or background objects) that people tag. We found that for European Americans, the first tag was more likely assigned to the main objects than that by Chinese; but for Chinese, the first tag was more likely assigned to the overall description or relations between objects in the images. The findings had significant implications for designing cultural-sensitive tools to facilitate the tagging and search process of digital media, as well as for developing data-mining tools that identify user profiles based on their tagging patterns and cultural origins.

Author Keywords

Cultural difference, annotation, tagging, image tagging, perception, attention

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors, Experimentation

INTRODUCTION

Innovations in technology have fostered a rapid growth in both personal and public digital media collections by providing a large variety of easy-to-use capturing devices and low-cost storage devices. A growing population begins to realize the benefits brought by semantic interpretations and annotations (e.g., tags) on searching tasks from these enormous collections of digital media [6, 8]. As a result, researchers in human-computer interaction (HCI) have been actively investigating human tagging behavior. To further ease the tagging process and motivate tagging behavior of

digital media users [3], a number of desktop based tools [2, 9] and online media sharing communities such as Flickr and Youtube, have been developed.

Studies have been conducted to better understand *why* and *how* people create tags to digital media such as images [6]. However, we think that the content of tags (*what*) created by users are also important for two reasons. First, one way to make the tagging process easier is to provide algorithms for partially automated tagging such as suggesting tags for digital media [10]. The more the suggestions resemble the actual tags created by human users, the easier the tagging process will be. Second, better tag-based search algorithms can provide a more visible and direct association between the effort of creating tags and the benefit of searching from tags, thereby better motivate users to create tags [1].

Another important aspect is the cultural origins of taggers. Research in perception and cognition suggest that people in different cultures allocate attention differently when viewing images and animations. Westerners tend to focus on main objects and pay less attention to background and contextual information, whereas Easterners have a holistic way of perceiving images, in the sense that they tend to equally spread attention more evenly over different parts of an image [7]. For example, Westerners mention the active focal object more often and describe the inert background objects and overall context less often than Easterners in their first sentence when describing an animation [4]. Westerners are also more likely to detect changes in main objects and less likely to detect contextual changes than Easterners [5]. Based on the assumption that users' tagging process is implicitly tied to their basic attentional processes, we hypothesize that culturally different attentional patterns may predict cultural differences in the order of tags people create for digital media. In other words, *we hypothesize that people from different cultures may exhibit different tagging patterns*.

In the current study, we compared the content of tags created by European Americans and Chinese for a same set of digital images. We examined the extent to which people in the two cultures assign tags to describe different parts (e.g., the main object, the background object, and the contextual information) of an image. We also examined the association between the sequential order and the content of tags. We hypothesize that compared to Chinese people European Americans will be more likely to assign tags to the main objects in earlier tags; but Chinese will be more likely to

assign tags to the background/contextual objects than European Americans.

METHOD
A between-subject quasi-experiment design was employed in the current study. Participants were recruited from two cultural groups. All participants were presented the same materials and experienced the same procedure. Participants' responses were compared between the two cultural groups.

Participants
Twenty-one European Americans (12 female) and 23 Chinese (15 female) were recruited from a university community to participate in the study. All European Americans were born in the United States and English was their first language except for one, who was born in France with French as the first language. None of the European Americans had stayed in an Eastern country for more than 9 months. All Chinese participants had stayed in the US for less than 15 months, with the average time spent in the US being 3.17 months. The two groups were similar in age ($M = 21.71$ years, $SD = 3.33$ for European Americans, $M = 22.38$ years, $SD = 2.35$ for Chinese), and education level (most had some college education and some had graduated from college). For both cultural groups, participants were familiar with searching for images on the internet and had some experience in tagging images on websites such as Facebook and Flickr.

Experimental Materials
Sixty digital images were selected from search results using public search engines such as Google. Criteria to select images were 1) photo of real-life objects; 2) there is no language or cultural iconic content in the image; 3) the image contained at least one clear foreground main object and a number of distinguishable background objects; 4) the main object belonged to one of the three categories: human, animal, and still objects; and 4) for a good proportion of the

Figure 1. Sample images in each category

images, there were similarities in either the main objects or the background objects, so that the participants would be invoked to create multiple tags to help distinguish one image from another. There were three groups of digital images, each with 20 images portraying a main object in one of the 3 categories mentioned above. A sample image in each category is provided in Figure 1.

Procedure
The procedure was the same for all participants, except that the experiment was conducted in English for European Americans and in Chinese for Chinese. We made sure that all participants understood the process of creating tags to images and the purpose of creating the tags. Participants were given the following instructions:

> "The purpose to create these tags is to make it easier to search for a particular image in the future. Please imagine that weeks or months later, when you come back to look for a certain image, you can only use tags as your searching cues. The tags you have created should be able to help you find the correct image faster."

Each participant saw one image at a time, at a random order. For each image, participants were asked to create at least one, at most 10 tags to describe the image using a web browser. Participants were also asked to create short, single word tags rather than long phrases or sentences. European Americans created all tags in English, and Chinese participants created all tags in Chinese. After tagging all images, participants were presented the same set of 60 images again, one at a time, in a random order, together with the tags they just created. Participants were then asked to select the tags they believe describe the main object of the image. For each image, participants were allowed to select at most 3 tags.

Coding Procedure
Four steps were taken to code tags into categories. First, three researchers (one from each culture, and one bi-cultural researcher who has spent more than 6 years in each of the two counties) went through all images and agreed on which part is the main object. Second, a coding scheme (Table 1) was developed and the coders were trained. Third, a Chinese coder did the coding for all Chinese tags and 2 European American coders coded the English tags. The bi-cultural researcher coded at least 10% of tags in each language to ensure that all coders have sufficient agreement with the bi-

Coding Category		Coding Criterion	Examples
Uncodable		The tag was not a recognizable word or the coder could not associate it to any part of the image	
Foreground Main Object	Name	The tag described the foreground main objects that were agreed on by the three researchers in the first step of the coding procedure	boy
	Property		cute
	Behavior		running
Background Object	Name	The tag described any of the objects that were not considered the foreground main object decided in the first step of the coding procedure	boats
	Property		white
	Behavior		sailing
Overall Description		The tag described overall features of the whole image, e.g., place, environment, event, time, emotion conveyed by the image, and photography technical terms	city, sunny, party, fall joyful, overexposure
Relationship		The tag described a relationship between two or more objects in the image	boy-in-park, above

Table 1. Coding scheme for the tags created by participants.

cultural coder. Inter-coder agreements ranged from 81.1% to 87.2%. Lastly, coders discussed the disagreements and reached a consensus on them.

RESULTS

Due to technical difficulties, data from one Chinese participant was excluded from the analysis. The average number of tags participants created and selected as describing the main object for each image category is presented in Table 2. On average, European Americans created more tags and selected more tags as describing the main object than Chinese. The two cultural groups selected a similar proportion of tags as describing the main object.

Image Category	American		Chinese	
	Tags	**Object**	**Tags**	**Object**
Animal	5.27 (2.54)	2.10 (.57)	3.96 (1.72)	1.61 (.63)
Human	5.42 (2.44)	2.08 (.61)	4.17 (1.81)	1.64 (.75)
Still Object	5.34 (2.44)	1.96 (.60)	4.01 (1.62)	1.65 (.62)
Total	5.35 (2.46)	2.05 (.57)	4.05 (1.70)	1.63 (.66)

Table 2. Average number of tags created and selected as main object tags (SDs are provided in parenthesis)

Cultural difference in Main/Background Object Tagging

The major investigation of the current study is to look at whether participants from the two cultural groups tag the main versus background objects differently. We took sequential position of tags into account based on the assumption that tags created earlier are usually associated with the salient parts of an image and/or considered more important in describing the image by participants [5]. Since not all participants created more than 3 tags for each image, the result from the 3rd to 10th tag was combined in the analysis. All analysis in this section generated similar patterns of results in the three image categories. Therefore, only the combined results across image categories are reported.

Participants' selection of tags

The percentage of tags participants selected as describing the main object was computed for each tag position for each participant across all 60 images. The mean percentages of tags selected by participants as describing the main object are presented in Figure 2(a). Mixed-design analysis of variance (ANOVA) with culture as between-subject variable and tag
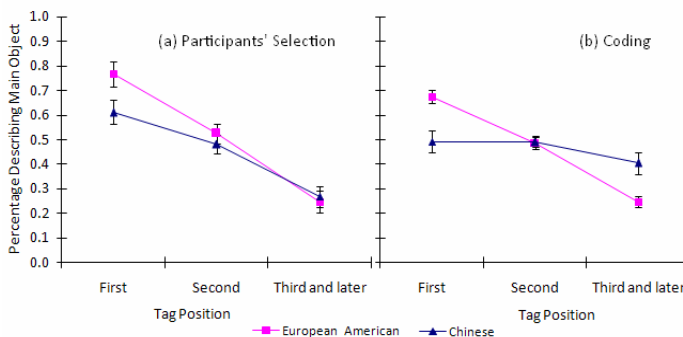

Figure 2. Percentage of tags describing the main objects for each tag sequential position

position as within-subject variable revealed an interactive effect between culture and tag position ($F_{2, 84} = 122.73$, $p < .01$) and a main effect of tag position ($F_{1, 42} = 5.37$, $p < .01$). Although the proportion of tags selected as describing the main object decreased along tag positions in both cultures (all ps $< .01$), consistent with our hypothesis that main objects were more salient to European Americans, they selected more tags as describing the main object than Chinese ($t_{42} = 2.21$, $p < .05$) in the most important position (i.e., the first ones) than Chinese.

Coding of tags

It is possible that the cultural difference in the selection of tags is caused by different definition of main objects by the two cultural groups. Thus, an objective coding scheme was developed to code the tags into different categories to ensure an equal criterion for selecting the main object across *both groups*. If a similar result pattern was found, this alternative explanation can be ruled out.

Based on the coding scheme, percentage of tags coded as describing the main objects was computed by summing up the three subcategories (name, property, and behavior) describing the main object for each tag position across all 60 images for each participant. The mean percentages of tags describing the main object are presented in Figure 2(b). Similar to the results with participants' selections, mixed-design ANOVA revealed an interactive effect between culture and tag position ($F_{2, 84} = 11.39$, $p < .01$) and a main effect of tag position ($F_{1, 42} = 26.79$, $p < .01$). Further analysis suggested that European Americans selected more tags as describing the main object than Chinese for the first position ($t_{42} = 3.39$, $p < .01$) and less tags for the third position ($t_{42} = 2.99$, $p < .01$). Moreover, the proportion of tags selected as describing the main object decreased along tag positions only for European Americans (all ps $< .01$).

The results from the coding of tags yielded a similar and even stronger effect in the same direction in our hypothesis. That is, the foreground main objects appeared to be more salient to European Americans as they described the main objects in earlier tags rather than later ones, whereas Chinese tended to describe the foreground and background objects equally likely for each tag position.

A Closer Look at the First Tags

The percentage of the first tags coded into each category in the coding scheme for each culture is presented in Table 3.

Coding Category		**American**	**Chinese**
Uncodable		**1.43**	**0.36**
Main Object	Name	**62.14**	**41.98**
	Property/Behavior	5.32	7.33
Background Object	Name	12.85	13.58
	Property/Behavior	0.71	1.23
Overall Description		**16.75**	**30.14**
Relationship		**0.79**	**5.37**

Table 3. Percentage of the 1st tags coded into each category

Pearson Chi-square (χ^2_6=159.72, p<.01) suggested between-culture difference in some of the categories, as flagged out in bold (standardized residual larger than 2.0). European Americans tagged names of the main objects more often in their first tags whereas Chinese started with tagging overall properties of the images and relationships between objects in their first tags. This pattern again provided support to the main hypothesis that European Americans tended to pay attention to and tag the main object in their first tag, while Chinese tended to pay attention to and tag the overall description and/or the relationship embedded in the image in their first tags.

We were also interested in whether image categories had an effect on how likely participants from the two groups tag the main object in their first tags. The percentages of the first tags that described the main object for each image category are presented in Table 4 for both participants' selections and the codings. Mixed-design ANOVA found no interactive effect. Culture had a main effect ($F_{1, 42} = 374.06$, $p < .01$ for participant selections; $F_{1, 42} = 466.27$, $p < .01$, for codings). That is, European Americans tagged the main objects in their first tags more often than Chinese in all three categories. Image category also had a main effect ($F_{2, 84} = 28.36$, $p < .01$ for participant selections; $F_{2, 84} = 92.43$, $p < .01$, for codings). For participants' selections, main objects were tagged more often in the animal category than the other two (ps < .01). For the coding results, the percentage of tags describing main objects was highest in animal category and lowest in the category of still objects. This is also consistent with previous research results that suggested that active objects are usually more salient than inert objects across cultures [4].

Image Category	Participant Selection		Coding	
	American	**Chinese**	**American**	**Chinese**
Animal	87.62	68.27	82.14	58.92
Human	73.10	58.50	69.29	53.49
Still Object	69.76	56.53	50.95	35.00

Table 4. Percentage of first tags describing main object

CONCLUSION AND DISCUSSION

This study examined the cultural difference in image tagging behavior. Participants' selections and researchers' coding provided convergent evidence in support of our hypothesis. A greater proportion of the first tags created by European Americans described foreground main objects than those created by Chinese participants, who were more likely to tag the overall properties of the images.

The findings provided implications on culturally sensitive design of tools that utilize user-generated tags as indices for image searching. Our findings would suggest that when using tags created by European Americans as search indices, one would more likely obtain images that were tagged based on the main objects; but for tags created by Chinese, the same search engine would return images that were tagged based on their overall descriptions. In other words, the same search engine would return different sets of images if no cultural-sensitive weightings were used. Future research is needed to understand whether users from different cultures would also exhibit similar cultural differences in the keywords that they used when they search for images. Results would be useful for design of intelligent image search tools that adjust for the cultural difference that we observed.

We believe that current results will also inform design of data-mining tools to identify different communities of users based on the similar patterns of image tagging behavior. For example, for the same image, users who assign tags in a particular sequence may imply that they are more likely to belong to a particular culture. Cultural-sensitive algorithms can then be designed to better assist their future tagging and search tasks. Of course this will require further research to validate the extent to which this also applies to cultural groups other than European Americans and Chinese.

REFERENCES

1. Dong, W., and Fu, W. T. Toward a Cultural-Sensitive Image Tagging Interface. In *Proc. IUI 2010*, ACM Press (2010).

2. Girgensohn, A., Adcock, J., Cooper, M., Foote, J. and Wilcox, L. Simplifying the Management of Large Photo Collections. In *Proc. INTERACT 2003*, IOS Press (2003), 196-203.

3. Kustanowitz, J. and Shneiderman, B. Motivating Annotation for Personal Digital Photo Libraries: Lowering Barriers while Raising Incentives. *Tech. Report HCIL-2004-18*, U. Maryland, 2005.

4. Masuda, T. and Nisbett, R. E. Attending Holistically Versus Analytically: Comparing the Context Sensitivity of Japanese and Americans. In *Journal of Personality and Social Psychology, 81*, 5 (2001), 922-934.

5. Masuda, T. and Nisbett, R. E. Culture and Change Blindness. In *Cognitive Science: A Multidisciplinary Journal, 30*, 2 (2006), 381-399.

6. Morgan, A. and Naaman, M. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proc. CHI '07*, ACM Press (2007), 971-980.

7. Nisbett, R. E., Peng, K., Choi, I., and Norenzayan, A. Culture and Systems of Thought: Holistic vs. Analytic Cognition. In *Psyc. Review, 108* (2001), 291-310.

8. Rodden, K. and Wood, K. How Do People Manage Their Digital Photographs? In *Proc. CHI 2003*, ACM Press (2003), 409-416.

9. Shneiderman, B. and Kang, H. Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In *Proc. InfoVis 2000*, IEEE (2000), 88-95.

10. Wenyin,L., Dumais, S., Sun, Y., Zhang, H., Czerwinski, M., and Field, B. Semi-automatic image annotation. In *Proc. INTERACT 2001*, IOS Press (2001), 326-333.