

Visual-Effect Dictionary for Converting Words into Visual Images

Shogo Hirai^(✉) and Kaoru Sumi

Future University Hakodate, 116-2 Kameda, Hakodate-shi, Hokkaido, Japan
g2117041@fun.ac.jp, kaoru.sumi@acm.org

Abstract. This paper describes a visual effect dictionary that can visually express images of words used as modifiers, such as adjectives, onomatopoeia by adding 3D objects with a visual effect. This visual effect dictionary links words and visual effects for the target objects. The same modifier can have different meanings depending on the target object thus, an intelligent algorithm to select the visual effects is required. This dictionary can be used as communication support for images that are difficult to express only by modifier words such as adjectives, onomatopoeia. When multiple regression analysis was conducted in the questionnaire results, it was found that there are appropriate visual effects and inadequate visual effects depending on nouns.

Keywords: Visual effect · 3D object · Image words

1 Introduction

In this research, we have developed a visual effect dictionary that can visually express images of words as modifiers such as adjectives, onomatopoeias as visual effects for 3D objects. Onomatopoeia is a word that phonetically imitates, resembles or suggests the sound. This visual effect dictionary is a dictionary that links a word and a visual effect of the target object. Development of visual effect dictionary is difficult because the same modifier could have different meanings depending on the type of the target object; thus, an intelligent algorithm of selecting visual effects is required.

WordsEye [1] is a system in which “a yellow elephant” is displayed as a 3D object when text such as “Yellow elephant” is inputted. Our system is different from that in that it shows visual effects by specializing on modifiers. Tanaka et al. [2] developed Kairai which expresses audio information by animation using 3D objects. Kairai is a natural-language understanding system that emphasizes spoken words such as 3D space viewpoint. Therefore, Kairai expresses about distances and colors such as “far” and “blue” as adjectival expressions. Similarly, our research also expresses onomatopoeic words as adjectives, but differs from Kairai, our system is targeted for animation and effects. Anime de Blog is also available in which 3D animation is displayed by entering characters [3–5]. Anime de Blog creates 3D animation by matching the motion of a subject character to a verb. In our system, adjectives or onomatopoeias is defined as surfaces of things generated by textures, visual effects, and animation.

For developing the algorithms, we analyzed the impression of participants by questionnaires using subjects and examined the algorithm for developing the visual effect dictionary. We believe that this dictionary is useful for supporting communication via images when we express impressions which is difficult to express only by words.

2 Visual Effect Dictionary System that Visually Expresses Images of Words

A game development engine system that used Unity5 and C# was employed as a development language. In this system, the user provides audio information to the system. The system converts the given audio information into a visual effect and a 3D object. The speech recognition of this system uses the IBM Watson Speech to Text. The recognized speech information is classified into “modifier” and “noun” by morphological analysis using the NMeCab system. The classified part of speech is classified by the IBM Watson Natural Language Classifier(NLC). According to the classified information, the “modifier” is converted into a visual effect, and the “noun” is converted into a 3D object and displayed to the user.

The classification method uses NMeCab, which can be used in Unity. NMeCab is a morphological analysis engine of the .NET library. This system classifies sentences into nouns and adjectives.

The noun classification uses IBM Bluemix NLC. The NLC classifies the contents inputted by the user into one of the classes prepared by using a previously trained classifier. The NLC automatically classifies the noun information entered by the user into these four classes with a possibility. For example, apple is classified in a class called “fruit” which is prepared in advance.

In this system, we classified the modifiers into seven categories with reference to “*E de manabu Giongo · Gitaigo Card*” (Onomatopoeia-mimetic word card studied with pictures) [6]. We created “glowing”, “disturbed”, “moisture”, “dynamic,” and “floating” as visual representation with visual effect. The 3D-prepared objects are “fruit,” “food,” “ball,” and “car.” The visual effect and 3D object classification method will be explained later. This system converts classified adjectives, onomatopoeias, into associated visual effects. The visual effects are of three types, namely, “visual effect around a 3D object”, “visual effect modified texture”, and “Movement to a 3D object itself.”

The visual effect-modified texture expresses colors such as “red” and “blue”. An example of a visual effect around a 3D object expresses abstract words such as “delicious”, “beautiful.” In the movement to a 3D object itself, the coordinates of the 3D object are varied to adapt to the expressions such as “fast” and “slow”. Visual effects are created by imagining the aforementioned “glowing”, “disturbed”, “moisture,” “dynamic,” and “floating.” We searched synonyms using Japanese WordNet and added the test data.

Nouns are displayed by importing the 3D object from the Asset Store of Unity. Currently available 3D objects are of five types, namely, “fruit”, “warm food”, “cold food”, “car”, “ball”. First, we chose foods for many adjectives that we often use adjectives such as taste, tactile, and soon. In addition, we chose fruits whose characteristics

can be easily understood, such as hard and soft. Even the system can be used with other foods; thus, we investigated whether we can use the same visual effect. Therefore, we chose food, in addition to fruit. The classification for food alone is too broad; hence, we divided the food into warm and cold foods. We selected a “ball” as that is similar shape of “apple” and an example of other than food. In addition to the above four types, we prepared a “car” as a classification that is likely to reflect many animated visual effects.

3 Experiment

We will introduce from the aspects of clustering classification, and multiple regression analysis based on the questionnaire according to the common points of visual effect, modifier word, and 3D object. In this experiment, the factors and groups were analyzed using clustering classification in order to obtain common items and unknown groups from the questionnaire result. Then, we conducted multiple regression analysis to investigate the correlation when we perform evaluation using common sense knowledge of objects.

3.1 Purpose of the Experiment

The purpose of this experiment is to investigate whether an image of a word can be correctly reflected by the visual effect and the 3D object. In this research, we assume that by investigating visual effects and 3D objects, we can explore new classification method that can be understood using visual effect.

3.2 Experimental Method

This experiment was conducted employing 30 male and female University students over the age of 18 years in December 2016. The questionnaires asked whether the subject that read the sentences consisting of “modifier & noun” from the combined animation of “3D object & visual effect” in order to investigate whether the subject can recall the modifier from the 3D object and visual effect. The questionnaire was created using Google Form. The questionnaire generated three classification types, namely, “fruit,” “food” and “others.” In each item, the “fruit” class is composed of 25 videos and 43 questions, the “food” class is composed of 23 videos and 37 questions, and the “others” class is composed of 28 videos, 42 questions. The subjects viewed the videos checked and answered the questions. For example, a subject watched a movie showing a combination of the glowing effect and a 3D-object representation of an apple. Next, the video presented a question of whether the subject can recall the “delicious apple.” The subject selected appropriate answers from five levels of recollection, namely, “can recall,” “can somewhat recall,” “do not know,” “cannot very much recall,” “cannot recall.”

3.3 Experiment Results

In the case of a combination of “*Oishii*” (delicious) apple and “glowing” effect, 60.50% of the subjects selected “can recall” or “can somewhat recall.” The “glowing” effect represented a majority of the answers in both “*Oishii*” (delicious) and “*Atarashii*” (new) ratios of the apple recall. With regard to bananas and peaches, the proportion of “new” represented the majority of the answers. The majority answers of the two objects of “warm food” accounted for the proportion that can be recalled with the “steam” effect for both “*Oishii*” (delicious) and “warm” effects. Regarding the “cold food,” more subjects cannot recall either effect. The evaluation of “ball” was dominated by a majority of the proportion in which “*Atarashii*” (new) can be recalled when using the glow effect was used. Majority of the subjects accounted for the proportion in which “cars” and “buses” can recall “*Atarashii*” (new) with the “glowing” effect.

3.4 Relationship of Common Sense Knowledge of Objects

We investigated the modifier “*Atarashii*” (new) from the common sense knowledge of the object. The common sense knowledge of the object cited common sense knowledge and general items that appeared as a result of searching each noun using a meaning dictionary called Japanese Wordnet. We conducted a multiple regression analysis. As a result, the decision coefficient (R^2) was 0.989, which shows that it was significant at the 10% level. The significance of the regression equation as a whole was 0.00, which was significant at 1% level. The standardization factor that each item was significant at a significant probability of 0.01% of the warm food was 0.000. We found that the cold one was significant at 0.011 and food was significant at 0.034. The significant probability was at 0.1% level. we found that the normalization coefficient of the warm one was 0.932, and the positive correlation provided even greater influence. In other words, the combination of “steam” and “warm food” turned out to be appropriate to express the modifier “*Atarashii*” (new). The following is a summary of the “new” “steam” visual effects. (1) We focused on the common sense of objects. (2) The result showed a positive correlation among warm, and cold things and food as common sense of objects appropriate for “steam” visual effects. We also focused on “*Oishii*” (delicious) using the same visual effects as different modifiers, and performed multiple regression analysis. From the multiple regression analysis using the variable reduction method with the combination of “*Oishii*” (delicious) and glow, the common sense of all objects was excluded, and the significance probability could not be calculated. In other words, the data were not appropriate data for multiple regression analysis and no relationship was found. Finally, we observed the combination of “*Oishii*” (delicious) and steam. The coefficient of determination (R^2) was 0.997 and turned out to be significant at the 10% level. The significance of the regression equation as a whole was 0.00, which was significant at the 1% level. The standardization factor of each item was significant at a significance probability of 0.01% level of the “warm food.” The normalization coefficient also showed a positive correlation of 0.939. In other words, in expressing the modifier “*Oishii*” (delicious), we found that the combination of “steam” and “warm food” was appropriate.

4 Discussion

According to the questionnaire result, a majority of subjects recalled “*Oishii* (delicious) apple” with the “glowing” effect and the “apple”. A majority of subjects recalled “*Atarashii* (new)” with the “glowing” effect and the “fruit”. However, a majority of subjects didn’t recall “*Atarashii* (new)” with the “glowing” effect and the “hot food” or “cold food”. Therefore, only the “glowing” effect of the fruits in the food turned out to represent the adjective “new.”

A majority of subjects recalled delicious, new and warm foods with the “steam” effect. In other words, “steam” can be remembered as delicious for expressing a new warm food.

As the multiple regression analysis, it was suggested that “hot food & steam” had a strong influence on the questionnaire result recall of “*Atarashii* (new)”. Further, we found that the combination of “warm food & steam” was appropriate for the modifier “*Oishii*(delicious)”. This analysis suggested that appropriate or inappropriate combinations are available for the visual effects, modifiers, and 3D objects. The modifiers are related to the common sense knowledge of 3D objects for visualizing. By examining this relationship in more detail, nouns can be classified into multiple classes, such as common sense knowledge of 3D objects. Thus, we could be concluded that we can visualize a modifier with multiple meanings by using common sense knowledge of the object. As a future work, we will increase the number of questionnaires, visual effects to use and nouns (Fig. 1).

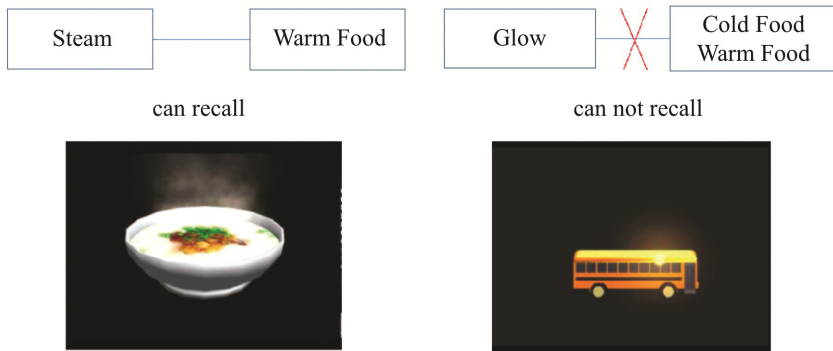


Fig. 1. Analysis summary

5 Conclusion

We have introduced the development of a dictionary that shows visual effects of adjectives, onomatopoeias that corresponding to 3D objects. According to the questionnaire result, it was suggested that “delicious” is easily recalled by giving a steam effect to warm food. Thus, by investigating from the common sense knowledge of the object, we

found that it is possible to provide an appropriate visual effect to the modifier using the system.

References

1. Bob, C., Richard, S.: WordsEye: An automatic text-to-scene conversion system. In: Proceedings of the 28th Annual Conference on Computer Graphics, SIGGRAPH 2001. ACM, Los Angeles (2001)
2. Hozumi, T., Takenobu, T., Yusuke, S.: Animated agents capable of understanding natural language and performing actions. In: Prendinger, H., Ishizuka, M. (eds.) *Life-Like Characters*, pp. 163–187. Springer, Heidelberg (2004)
3. Sumi, K.: Anime de Blog: Animation CGM for content distribution. In: Proceedings of International Conference on Advances in Computer Entertainment Technology (ACE2008), SIGCHI, pp. 187–190. ACM (2008)
4. Sumi, K.: Animation-Based Interactive Storytelling System. In: Spierling, Ulrike, Szilas, Nicolas (eds.) *ICIDS 2008*. LNCS, vol. 5334, pp. 48–50. Springer, Heidelberg (2008). doi: [10.1007/978-3-540-89454-4_8](https://doi.org/10.1007/978-3-540-89454-4_8)
5. Kaoru, S.: Capturing common sense knowledge via story generation. In: *Common Sense and Intelligent User Interfaces 2009: Story Understanding and Generation for Context-Aware Interface Design*, 2009 International Conference on Intelligent User Interfaces (IUI2009), SIGCHI. ACM (2009)
6. Tomikawa, K.: *Onomatopoeia/mimetic word card studied with pictures*, 3anet (1997)