

How to Teach Young Kids New Concepts with Interactive Videos and Visual Recognition

Quan H. To, Ba-Huu Tran^(✉), and Minh-Triet Tran

Faculty of Information Technology, University of Science, VNU-HCM,
Ho Chi Minh City, Vietnam
{thquan,tbhuu}@apcs.vn, tmtriet@fit.hcmus.edu.vn

Abstract. Advances in modern computing technology have enabled a wide variety of applications in many areas. Realizing that mobile devices such as smartphones and tablets are great learning platforms for educating young children, the authors create an interactive system that helps young kids learn new concepts via videos. By allowing kids to choose objects in the video that they are watching, the system provides relevant information about those objects. It has the effect of enhancing young kids' understanding about new concepts in the video. The system employs Cross-based Local Multipoint Filtering (CLMF) for object selection and Support Vector Machine (SVM) for recognition. Experimental results show that our system has achieved 95 % recognition accuracy.

Keywords: Learning platform · Educational technology · Interactive system · Visual recognition · Human computer interaction

1 Introduction

Getting to know the world around them is essential to young kids' mental development, especially in the early years [9]. During this critical period, young kids learn new concepts (such as toys, animals, flowers...) in many ways. They observe the scenery, read picture books and listen to stories told by their parents, watch videos... As a result, there are various means of education to enrich kids' understanding of the world. One of the most popular means to educate kids is by using printed or digital books, because books can capture children's attention with beautiful pictures. The authors explore a way to enhance the interaction between young kids and the teaching materials, as well as to get them more engaged, thereby helping kids learn more effectively.

In this paper, we propose an innovative and interactive system for young kids to learn new concepts on mobile devices, such as tablets and smartphones. This system enables young kids to interact with the videos that show different concepts and objects in the real world. Videos can easily capture children's attention via motions, sounds, and colors. Another advantage of using videos is that it can teach kids about the context of each concept; e.g. animals live in their natural habitats, food is made in the kitchen...

When kids watch a video, they see many interesting things and would want to know more about them. To learn about a certain object in the video, they simply draw a loose boundary around that object. Our system would automatically highlight that object

around its natural border, instead of using the traditional rectangular bounding box, which may include the background and minor artifacts. The background and minor artifacts may not be what kids want to learn about and could confuse or distract them. The system accomplishes this task by using the CLMF algorithm. CLMF helps identify the main object appearing in the selected area. As a result, the system can get an appropriate shape of that object. This allows young kids to know the real shape of the object that they wish to learn about.

Next, the system employs recognition algorithm to recognize the object bounded by its natural border. By removing the background and irrelevant artifacts that might appear near the object of interest, the authors can enhance the accuracy of object recognition. After that, the system displays the name of the highlighted object. It can also provide relevant information such as songs, stories, videos... about that object.

The main contributions of our paper are as follow:

- First, we propose an idea and implement the system on mobile devices to teach young kids new concepts via interacting with objects in video.
- Second, we propose using CLMF to determine the natural border corresponding to the actual shape of an object (animals, toys, trees...) that appears in the region of interest. This way, young kids can know exactly the actual shape of the object.
- Third, by removing irrelevant information in the region selected by young kids and keeping only the main object itself, we can enhance the accuracy of object recognition. Therefore, we can provide relevant information about that object to the young kids.

2 Related Works

2.1 Cross-Based Local Multipoint Filtering (CLMF) [1]

Edge-aware filtering (EAF) techniques are widely used in many areas of image processing. Two of the most popular EAF processes are bilateral filter (BF) [2] and guided filter (GF) [3]. However, both of them have their own weaknesses. For example, BF produces staircase effect and gradient reversal artifacts while GF generates undesired fuzzy boundary. CLMF could overcome the weaknesses of both BF and GF. Therefore it emerged as a better alternative because of its fast performance and high quality.

2.2 Bag-of-Words Model

Usually, hand-crafted descriptors such as SIFT [4], SURF [5] are used to represent object's local features. The extracted features are then clustered to build the visual vocabulary. In the Bag-of-Words model, objects are represented by a histogram of quantized local features, i.e. the frequencies of the visual words [6, 7]. After all the description vectors have been computed, they serve as the training examples for the Support Vector Machine (SVM). Finally, the SVM model returns the label of the object we want to classify. The Bag-of-Words model has been used extensively in computer vision to solve a wide variety of recognition problems with remarkable results.

3 Proposed System

3.1 Overview

Figure 1 illustrates the overview of the whole system. First, young kids simply open the application on a mobile device and choose an arbitrary video that they want to watch. The video shows wonderful things that exist in this world. Naturally, the kids would love to know more information about various objects that they see in the video. We enhance the interactivity of the video by letting the kids select any object within that video. To do that, they simply use their fingers to draw a free closed shape around that object. Our system will automatically identify the region in which the kids are interested; and then show a natural border around the interested object. With this, the young kids would be excited because the object of interest is highlighted with a natural border that closely corresponds to its actual shape in reality. Additionally, the object is extracted without unnecessary background and irrelevant artifacts, i.e. minor objects that is not of the kids' interest. The system then classifies the chosen object using an object classification module. The output of this module is the label of the class to which this object belongs. Finally, the system provide appropriate augmented information on screen. The augmented information could be in different formats; including texts, 3D models, video clips, audio clips, web pages...

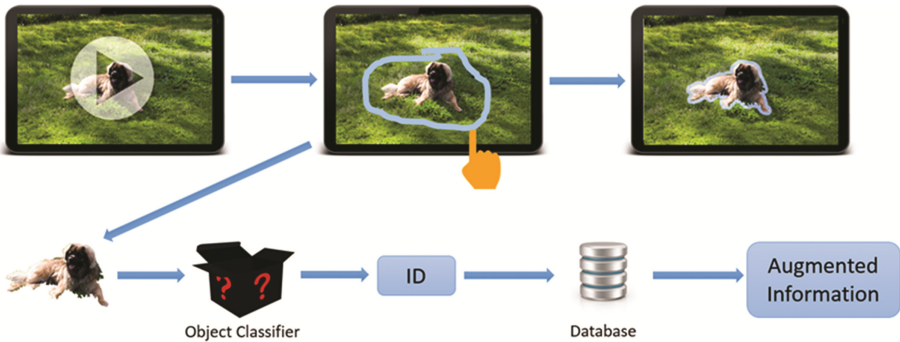


Fig. 1. Overview of our proposed system

3.2 Determining the Natural Border of an Object

Figure 2 represents how the system can approximate the natural border of an object. After the kids have drawn the free closed boundary around the object, the area inside this boundary forms the region of interest. The region of interest is used to create a binary mask. However, this binary mask is not well-aligned with the natural border of the object that the kids choose. As a result, we apply CLMF to transform the binary mask. CLMF uses the original image to guide the process of turning the mask into a shape that aligns more closely with the chosen object. After this process, the binary mask now looks like the object that the kids are interested in. Finally, the system uses the resulting mask to highlight the object in the frame, as well as to extract the object for classification.

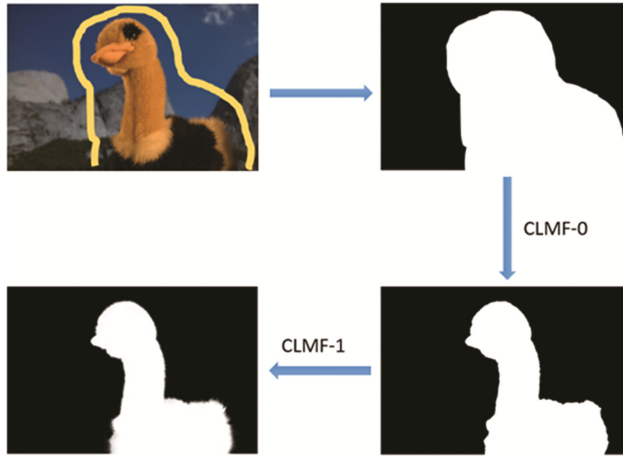


Fig. 2. Approximating object's natural border with pre-drawn boundary and CLMF

3.3 Object Classification with Support Vector Machine

Dense-SIFT features and Bag-of-Words model are used to represent the objects produced by CLMF algorithm. The authors employ multi-class Support Vector Machine (SVM) for object classification. SVM is a popular supervised learning model. The set of training examples is fed into the SVM. Each example is represented as a point in space and is marked as either positive or negative. SVM aims to solve the problem of finding a hyperplane that separates the positive and negative examples by a margin that is as wide as possible. New examples are then classified as belonging to one category or the other. By combining multiple two-class SVMs, we obtain a multi-class SVM that can recognize many types of object [8].

4 Experiment

4.1 Experiment Setup

To demonstrate the use of CLMF algorithm, the authors perform experiment on a set of five images with pre-drawn boundary. CLMF would determine the border of the object so that it is as close to the shape of the object as possible.

In the next part of the experiment, five short videos about several objects are selected for visual recognition. In each video, 10 frames are sampled and divided equally for the training and testing sets. There are 5 object classes, which are giraffe, cheetah, elephant, squirrel, and bird. The training set is fed into the SVM as examples. The authors then measure recognition accuracy on the testing set.

4.2 Experiment Results

Here is the original images and corresponding results when applying CLMF:

In the visual recognition experiment, the authors set the following parameters for the multi-class SVM: $C = 0.2$ and $\text{Gamma} = 0.15$. The system achieves a recognition accuracy of **95 %** (Fig. 3).

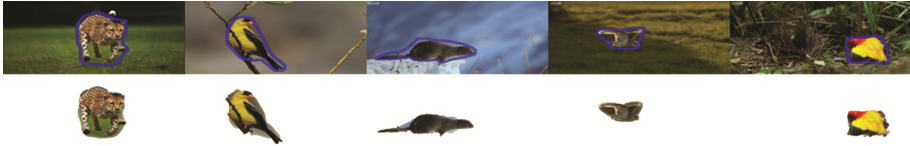


Fig. 3. Results of CLMF algorithm

5 Conclusion

In this paper, we propose a system to help kids learn new concepts via interactive videos. Young kids can draw a boundary around an object in the video to know more about it. By highlighting the object around its natural border, we can increase both kids' perception of the object; as well as the accuracy of object recognition. The system aims to educate and enrich young children's understanding of the world around them.

Acknowledgement. This research is supported by research funding from Advanced Program in Computer Science, University of Science, Vietnam National University - Ho Chi Minh City.

References

1. Lu, J., Shi, K., Min, D., Lin, L., Do, M.N.: Cross-based local multipoint filtering. In: CVPR (2012)
2. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV (1998)
3. He, K., Sun, J., Tang, X.: Guided image filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 1–14. Springer, Heidelberg (2010)
4. Lowe, D.: Distinctive image features from scale-invariant keypoints. *IJCV* **60**, 91–110 (2014)
5. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
6. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: International Conference on Computer Vision (2003)
7. Chatfield, K., Lempitsky, V., Vedaldi, A., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: BMVC (2011)
8. Weston, J., Watkins, C.: Support vector machines for multi-class pattern recognition. In: ESANN (1999)
9. Shonkoff, P.: From neurons to neighborhoods: the science of early childhood development (2000)