



Marker-free Object Recognition on Tabletops: An Exploratory Study

Frédéric Rayar, Armand Renaudeau

► To cite this version:

Frédéric Rayar, Armand Renaudeau. Marker-free Object Recognition on Tabletops: An Exploratory Study. 10th ACM International Conference on Interactive Tabletops and Surfaces, Nov 2015, Funchal, Madeira, Portugal. hal-01223488

HAL Id: hal-01223488

<https://hal.science/hal-01223488>

Submitted on 13 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Marker-free Object Recognition on Tabletops: An Exploratory Study

Frédéric Rayar

Laboratoire Informatique
Université François Rabelais
Tours, France
frederic.rayar@univ-tours.fr

Armand Renaudeau

Polytech Tours
Université François Rabelais
Tours, France
renaudeau@etu.univ-tours.fr

Abstract

In the last decade, object recognition placed on tabletop has been advertised as a near future ability on next generation tabletops by several major information technology companies. Markers-based recognition seems the straightforward solution, but one can easily understand that this may not be a lasting solution for real word scenarios. In this paper, we study the marker-free object recognition ability of a commonly used tabletop. This tabletop does not rely on external cameras and thus fits perfectly a future connected house scenario. A proof-of-concept system has been developed and experimented on various objects. Based on this study, we raise some conclusions and perspectives on both software and hardware levels.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User interfaces.

Author Keywords

Tabletop; Object Recognition; Marker-free

Introduction

Since almost ten years, major information technology companies release videos of near future life. The imagine scenario in everyday life [2] and various fields, such as

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
ITS '15, November 15–18, 2015, Funchal, Portugal
ACM 978-1-4503-3899-8/15/11.
<http://dx.doi.org/10.1145/2817721.2823508>

retail, banking, manufacturing, health or productivity. This near future, designed for the next decade, often presents tabletops and surfaces as common piece of furniture integrated in homes or offices. People will be able to interact with them only by placing various objects on them.

The objects are mostly electronic devices (smartphones, tablets), but also more common objects such as keys, books, food. For the former category, connected objects paradigm solves the communication issue with protocol such as NFC, Bluetooth, or Wifi. For the latter, computer vision techniques can be used. We address the latter category in this study. Two approaches have been proposed: markers-based or marker-free objects recognition. A marker is a two-dimensional pattern that can carry information and label object, e.g. barcode used in the industry, QR codes or fiducial markers. Such markers can be recognized in a camera feed and a system can deduce the related objects.

The marker-free approach is based on computer vision techniques that recognized the objects themselves instead of the markers. We focus the study on this second approach. Indeed, while marker object recognition can be a straightforward solution for usage such as tabletop games, scenarios such as cooking table, or daily life objects recognition call for marker-free object recognition.

In this paper, we propose to perform a study on marker-free object recognition on a commonly used tabletop in the HCI community. State-of-the-art image-based object recognition techniques will be used to evaluate the performance of the tabletop on this specific task.

Related work

To the best of our knowledge, the object recognition on tabletop for non connected objects has not been widely studied.

The first paradigm that can be found in the literature is the use of external cameras pointing at the tabletop. The camera recognize objects placed on the tabletop and communicate the result of the recognition task. It is then up to the tabletop to display digital content. Image-based object recognition [5] is a well studied task in computer vision and machine learning communities. It has application in several field such as robotics or surveillance. This solution may work well, but the idea of external cameras all around the house might be an issue to inhabitants.

The second paradigm is camera-aware tabletops. Here the vision ability is embed within the tabletop itself. Several propositions have been made: first, usage of a camera underneath the table and eventually a projector to display digital content. For instance the Lumisight table [6]. Second, is the use of infrared lights and sensors, as in the Samsung SUR40 [1]. The infrared allows the tabletop to have a raw image of the objects placed on it. With this solution, the table can be open in the bottom (i.e. a table top and its legs). This latter approach seems a better fit for the use of several tabletops in homes.

In this paper, we study the recognition performance of the Samsung SUR40, without the usage of any external devices in the tabletop vicinity.

Prototype

Objects description

Each object has to be described by visual descriptors. One can find several types of such descriptor in the



Figure 1: Raw image of a white smartphone obtained with the PixelSense.

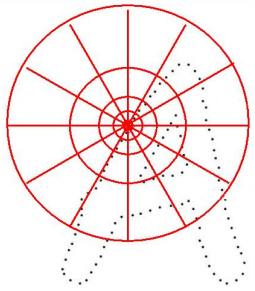


Figure 2: Shape context illustration.

literature, such as colors, textures or shapes. In our first experiments, SURF descriptors [3] have been used, but led to bad recognition rate. This is mainly due to the quality of the images obtained by the PixelSense 1. To fit the type and quality of our images, we have selected a shape based descriptor, namely Shape Context descriptor [4].

Shape context descriptor aims at capturing the relative distribution of points relatively to each point on the shape. It is computed as follows: (i) a set of point in the shape is extracted, (ii) for each point, a target is centered and the number of points in each cells is computed, (iii) the descriptor vector is computed using the previously computed histograms. Figure 2 illustrates the idea behind shape context. The matching between two shape context descriptors is done using a bipartite graph matching problem [4].

Objects recognition

We use here a classic supervised learning workflow (see Figure 3). First, the training phase is performed. A set of labeled objects to be recognized has been selected. These objects have been placed on the tabletop, and the relevant images have been obtained. Then, shape context descriptors have been computed (feature extraction) for each images, and stored in a database.

When a new object is placed on the tabletop (prediction phase), its shape context is computed, and matches against the database one. Using a nearest-neighbour classifier, the system find the most probable class of the new object: the object is recognized. If the matching does not have a good confidence rate, the system can conclude that the object have not been learnt.

As the complexity of finding the nearest match in the database is high (in $O(n^3)$), we have added a simple

dimensions based filter to reduce the number of shape context descriptor comparisons. Thus, we can perform real-time recognition of objects.

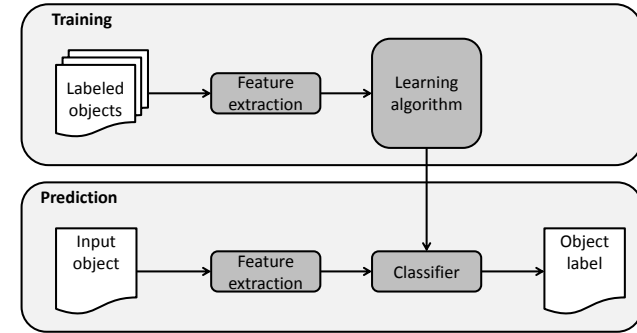


Figure 3: Supervised learning workflow.

Interface

An application has been developed on the PixelSense tabletop using C# language. Figure 4 illustrates the interface of the application, where four objects are recognised. The application can recognise as set of learned objects. When an object is correctly not recognised, the application allows the user to add it in the database. An administrator interface allows to validate or not the proposed new objects.

Experimental study

Seven distinct objects have been selected to assess the recognition capabilities: a 50 cent euro coin, our institute student card, our institute booklet, a business card, a rubber, a ruler and a white smartphone.

We have used the custom administrator interface to make the system learn these objects.

	Coin	Student card	Booklet	Business card	Rubber	Ruler	Smartphone
Recognised	20	20	20	16	18	19	19
Not recognised	0	0	0	0	2	1	1
Wrongly recognised	0	0	0	4	0	0	0
Recognition rate	100%	100%	100%	80%	90%	95%	95%

Table 1: Recognition rate of the seven selected objects.

Recognition rate

Each object has been placed 20 times. Table 1 shows the recognition rate obtained for the seven selected objects.

The system gives an average recognition of 94%, which is promising. Nevertheless, we have to put things into perspective: indeed, most of the selected objects have discriminant shapes. When this is not the case, the system may wrongly recognise some objects. For instance, the student card and the business card have more or less the same shape. That is why the business card has been wrongly recognised as a student card four times. This is due to a property of the shape context descriptor: it is invariant to scale and minor deformation. This phenomenon has been verified during a scenario where different coins (1,2,5,10,20,50 cent euro coins) have been used as objects to be recognised. The system failed most of the times by wrongly recognising another coin.

Recognition time

Another aspect that has to be experimented is the time that the system takes to recognise a given object. We have measured the amount of times taken by each of the following steps of our workflow: (i) image processing, (ii) dimensions based pruning, (iii) shape context descriptors matching and (iv) interface update. Table 2 presents the results of this experiment. One can see that the recognition could be done in a few seconds in average.

This is relevant for proof-of-concept purpose.

Nevertheless, even with some pruning based on the dimensions, the shape context matching step is still costly. Thus, the question of the relevance of the shape context descriptor raises with regards to the time computation.

Conclusion

In this paper, we have presented an exploratory study on the object recognition ability of tabletops, that have been sold as a near future ability on next generation tabletops. We focused our study on camera-aware tabletop that do not rely on external cameras, since they fit well a future connected house scenario.

An application has been developed as a proof-of-concept. It obtains relevant recognition rate for medium sized objects that have discriminant shapes. But due to the features selected to describe an object, wrong recognition can easily happen. Nevertheless, the recognition times obtained in the experiments is promising for real time recognition.

Based on this study, one can wonder about the possibility to improve such a system with better image processing and machine learning methods. Nevertheless, we also wonder about the possibility to conceive an innovative hardware solutions that will provide a good trade-off

	Coin	Student card	Booklet	Business card	Rubber	Ruler	Smartphone
Image processing	0	0	0	0	1	0	0
Dimensions pruning	0	0	0	0	1	0	0
Shape context matching	0	3	1	2	4	2	1
Interface update	0	0	0	0	1	0	0
Total	0	3	1	2	4	2	1

Table 2: Recognition times of the seven selected objects (in second).

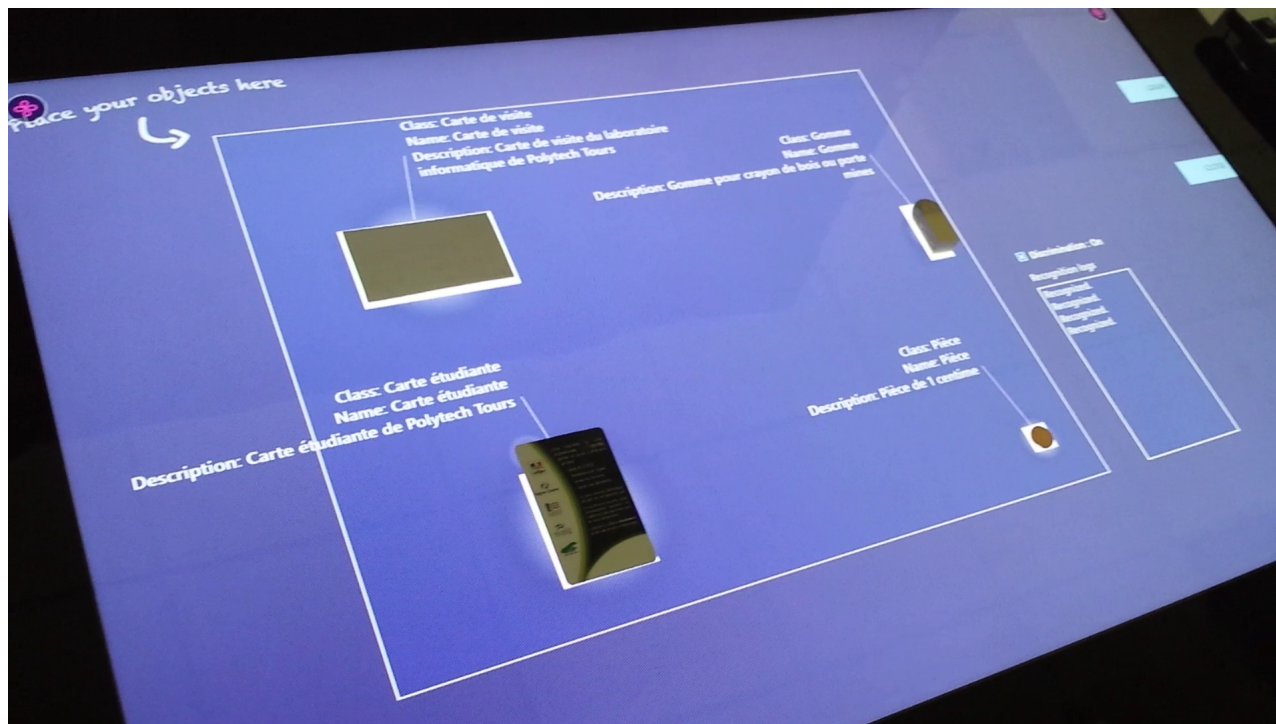


Figure 4: Interface of the tabletop application. Four objects have been placed and correctly recognised.

between the design simplicity, the robustness and the recognition performance.

Future works will try to handle real case scenarios such as the recognition of different coins (helpful for elders) or different student cards (registered collaborative work).

Acknowledgments

The authors wish to thank Ping Xiang for her work on the first prototype, and Pr. Jean-Yves Ramel for his guidance on this work. This work would not have been possible without the hardware provided by Polytech'Tours.

References

- [1] Microsoft PixelSense. <https://www.microsoft.com/en-us/pixelsense/pixelsense.aspx>, 2011.
- [2] Welcome To The Future. <https://www.youtube.com/watch?v=XyIvSIY0MTM>,

2014.

- [3] Bay, H., Tuytelaars, T., and Van Gool, L. Surf: Speeded up robust features. In *Computer Vision ECCV 2006*, vol. 3951 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, 404–417.
- [4] Belongie, S., Malik, J., and Puzicha, J. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, 4 (Apr 2002), 509–522.
- [5] Jain, R., Kasturi, R., and Schunck, B. G. *Machine Vision*. McGraw-Hill, Inc., New York, NY, USA, 1995.
- [6] Kakehi, Y., Iida, M., Naemura, T., Shirai, Y., Matsushita, M., and Ohguro, T. Lumisight table: an interactive view-dependent tabletop display. *Computer Graphics and Applications, IEEE* 25, 1 (Jan 2005), 48–53.