

Interactive Museum Guide

Herbert Bay, Beat Fasel and Luc Van Gool
Computer Vision Laboratory (BIWI)
ETH Zurich
Sternwartstr. 7
8092 Zurich, Switzerland
{bay, bfasel, vangool}@vision.ee.ethz.ch

ABSTRACT

In this paper, we describe the prototype of an interactive museum guide. It runs on a tablet PC that features a touchscreen, a webcam and a Bluetooth receiver. This guide recognises objects on display in museums based on images of the latter which are taken directly by the visitor. Furthermore, the computer can determine the visitor's location by receiving signals emitted from Bluetooth senders in the museum, so called BTnodes. This information is used to reduce the search space for the extraction of relevant objects. Hence, the recognition accuracy is increased and the search time reduced. Moreover, this information can be used to indicate the user's current location in the museum. The prototype has been demonstrated to visitors of the Swiss National Museum in Zurich.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces - Input devices and strategies; I.2.10 [Artificial Intelligence]: Vision and Scene Understanding; I.4.8 [Image Processing and Computer Vision]: Scene Analysis - Object recognition

Keywords

Object recognition, Interactive museum guide, Bluetooth

1. INTRODUCTION

Many museums present their exhibits in a rather passive and non-engaging way. The visitor has to scan a booklet in order to find some general information about the object. However, searching for information about object after object is quite tedious and the information found does not always cover the visitor's specific interests. One possibility of making exhibitions more attractive to the visitor is to improve their interaction with the guide. In this paper, we present an interactive museum guide which is able to automatically find and retrieve information about the objects of interest on a laptop-like device. Moreover, it provides further links and references allowing the visitor to browse comfortably on the Internet for an even broader description of the object.

Permission to make digital or hard copies of all or part 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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Smart Environments and Their Applications to Cultural Heritage.

UbiComp 2005, September 11-14, 2005, Tokyo, Japan.

1.1 Related Work

Recently, several approaches and methods have been proposed that allow visitors to interact with an automatic guide in a museum. Kusunoki et al. [7] proposed a system for children that uses a sensing board which can rapidly recognise types and locations of multiple objects. It creates an immersive environment by giving audio-visual feedback to the kids. Other approaches are robots that guide users through museums [4, 10]. However, such robots are difficult to adapt to different environments, and they are not appropriate for individual use. An interesting approach using hand-held devices, like mobile phones, was proposed by [5], but their recognition technique is limited to constant illumination.

1.2 Our Approach

We present an interactive, image-based museum guide that is invariant to changes in lighting, viewpoint, scale (zoom) and rotation. Our method was implemented on a tablet PC using a conventional USB webcam for image acquisition, see Figure 1. This hand-held device allows the visitor to simply take a picture of an object of interest from any position and is provided, almost immediately, with a detailed description of it.

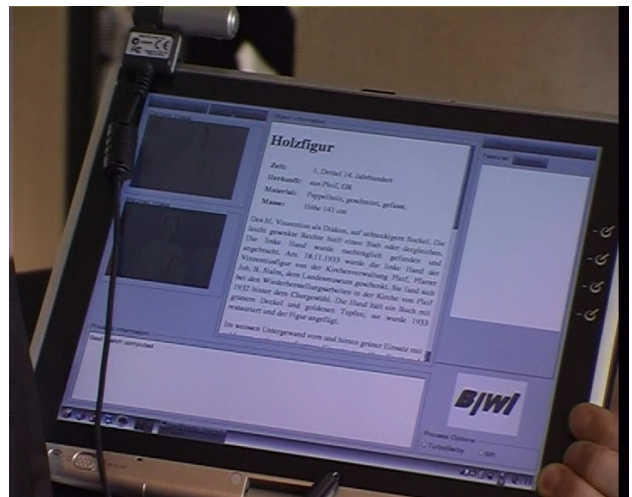


Figure 1: Tablet PC with the USB webcam fixed on the screen. The interface of the object recognition software is operated via a touchscreen.

Furthermore, this device can be extended to display location dependent information on a map, such as the closest emergency exits, the toilets or the direction to the next

coffee shop relative to the visitor’s position. Our museum guide neither imposes a predefined visiting order, nor the inconvenient task of scanning a vast database.

The museum guide has been shown to the public in the framework of the 150 years anniversary celebration of the Federal Institute of Technology (ETH) in Zurich, Switzerland. It was demonstrated in the Swiss National Museum Zurich. About 250 visitors took part in the demonstration in 20 guided tours of 10-15 persons each. The object descriptions were read by a synthetic computer voice, which enhanced the comfort of the guide.

2. METHOD

Our interactive museum guide contains two different modules. The first is an image-based object-recognition module, and the second consists of an automatic exposition room detector using Bluetooth. The combination of both techniques results in a robust and fast object recognition for large image databases.

2.1 Object Recognition

In order to retrieve the correct object, a database of images has to be established containing images of each object taken from different viewpoints. This fact assures a certain viewpoint independence of the guide and allows it to estimate the approximate direction from which the visitor took the picture. This information can be used as an extension for a more detailed, viewpoint dependent description. An example of a model image set can be seen in Figure 2.



Figure 2: Sample of model images and an input image (lower right image) of an object in the museum. Note the important differences in appearance between the model images and the input image. Also, scale and viewpoint of the input image differs from those of all the model images.

For each image, a set of interest points is computed and described by a scale and rotation invariant descriptor. For that task, we developed a fast SIFT [8] approximation using integral images. Our descriptor has the same number of dimensions (128), but is six times faster than SIFT, detects in average 15% more interest points, and shows similar performance. Due to the space limitation of this paper, it is unfortunately not possible to provide a detailed description of our approach. However, in the following we briefly mention the main difference to the SIFT descriptor.

In order to attain the important difference in speed, we use integral images as defined in [11]. The use of integral images enhances the speed for interest point detection and description. The entry of an integral image $I_{\Sigma}(\mathbf{x})$ at a

location $\mathbf{x} = (x, y)^{\top}$ represents the sum of all pixels in the base image I of a rectangular region formed by the origin and \mathbf{x} .

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j). \quad (1)$$

Once the integral image computed, it is easy to calculate the sum of the intensities of pixels over any upright, rectangular area.

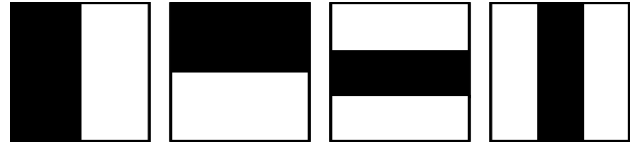


Figure 3: Haar-wavelet like patches approximating the first and second order derivative. The sum of intensities of the pixels lying in the black region is subtracted from the one in the white region

Integral images can be used to very quickly approximate the convolution kernels of the first and second order derivatives. The approximation consists in using Haar-wavelet like patches as illustrated in Figure 3. In this way, important accelerations can be achieved for the detection of Hessian-matrix based interest points as well as the required gradients for the SIFT descriptor.

In order to recognise the correct object from the database, we proceed as follows. The input image, taken by the user, is compared to all model images in the database by matching their respective interest points. The object figured on the model image with the highest number of matches with respect to the input image is chosen as the object the visitor is looking for.

The matching is carried out as follows. An interest point in the input image is compared to an interest point in the model image by calculating the Euclidean distance between their 128-dimensional descriptors. A matching pair is detected, if its distance is closer than 0.6 times the distance of the second nearest neighbour. This is a common robust matching strategy [2, 8, 9].

The average detection time for a database of 130 images of 22 objects is about 10 seconds. The reason for this relatively long recognition time lies in the fact that the feature description vector is high-dimensional. Furthermore, for every recognition step, an average number of 230 descriptor vectors for the input image have to be compared to about 30000 of such vectors in the database. However, the recognition time can be reduced by an order of magnitude by using the Best-Bin-First algorithm [3]. Another approach to get similar results at lower recognition time was proposed by [6], and uses PCA on patches of the gradient image around the interest points. However, both methods suffer of either a loss in quality, or a more time consuming descriptor evaluation. We are currently working on a fast matching alternative using integral images.

2.2 Automatic Room Detection

In every exposition room, one or more Bluetooth senders, also called BTnodes [1], see Figure 4, are positioned.

A BTnode is a versatile, autonomous wireless communication and computing platform based on a Bluetooth radio, a second low-power radio and a micro controller. Every BTnode covers a specific area of the museum and provides it with a localisation signal broadcasted at constant intervals. The signal received by the interactive museum guide is used for two purposes. First, the position of the visitor can be evaluated and displayed on a map. Moreover, as mentioned above, further location-dependent information may be retrieved. Second, as several images of an object are needed in order to robustly recognise it, the number of images in the database increases rapidly depending on the number of objects featured in the museum. This fact slows the object recognition process down drastically. Moreover, as more similar objects may enter the database, the accuracy of the recognition decreases. Classical object recognition methods would be computationally too expensive to get any result in time. To increase the matching speed, the search space is reduced to objects in the area close to the visitor. This area is defined with the signal of a BTnode. Hence, for a faster and more accurate object recognition, only objects situated in this area are considered as candidates.

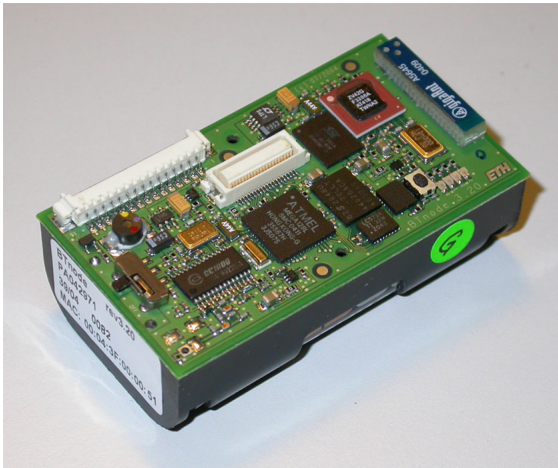


Figure 4: Image of a BTnode. These devices were placed in different exposition rooms of the museum. Each node broadcasts its identification number at regular time intervals.

2.3 Adding New Objects

Adding new objects to the database is easily accomplished. First, a few model images of the object have to be taken from different viewpoints with any kind of camera. The size of the image must be reduced to a conforming size in order to get a reasonable detection time without losing important details of the object. We chose 320×240 pixels. Second, interest points of the model images have to be detected and represented by our scale and rotation invariant descriptor. Finally, the model image names have to be indexed in a table in order to attribute the documentation to the figured object. Additionally, the number of the BTnode, covering the area where the object is located, has to be mentioned in the table.

3. APPLICATION

As soon as the computer receives the signal of a BTnode, it recognises the room in which it is located and selects the part of the database representing the objects in that same room. For the demonstration in the Swiss National Museum, we used only two of such BTnodes (see Figure 5), one in the entrance hall and the other in the first exposition room of the museum. Once the visitor passes the threshold to the entrance hall, the computer receives the signal of the first BTnode and *says* "Welcome to the Landesmuseum¹". As soon as the visitor enters the exposition room, the computer *says* "Exposition room" and launches automatically the object recognition application. The interface of the latter is shown in Figure 6.

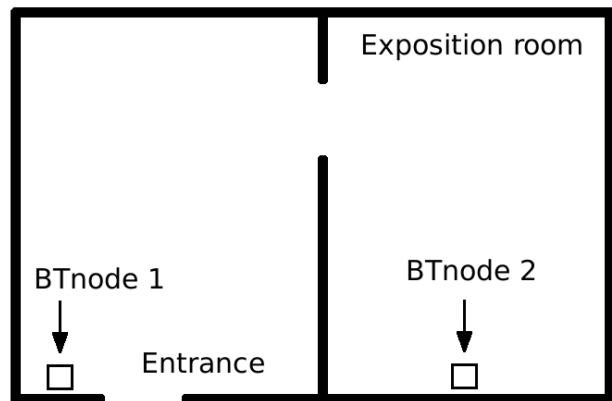


Figure 5: Schematic representation of the BTnode distribution in the museum.

When the user takes a picture of an exhibit, the computer displays, after a short computation time, the requested information in a browser window. Furthermore, the visitor can browse to some more specific information on the Intranet/Internet or to related objects that are currently exposed in the museum (e.g. made by the same artist). Also, the visitor has the option to have the description in the browser to be *read* by the computer via a text-to-speech synthesis engine.

4. RESULTS

Our interactive museum guide has been tested for 22 objects (130 database images) such as wooden statues, paintings, metal and stone items as well as coins and objects enclosed in glass cabinets which produce interfering reflections. The input images were taken from substantially different viewpoints under arbitrary scale (zoom) and rotation. We achieved an object recognition rate of about 80% for about 200 test images taken under various conditions. Furthermore, the recognition rate is affected by conflicts e.g. if two different objects were visible in the same input image.

This performance is quite promising, considering the fact that the interactive guide has to operate in an environment with varying conditions. It is robust to important natural lighting variations, such as different external weather conditions and daytime changes. Moreover, the results are not affected by artificial lighting or even changes in the colour

¹Landesmuseum means "National Museum"

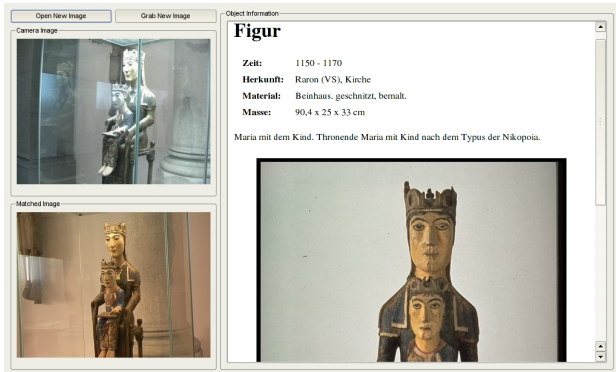


Figure 6: Interface of the object recognition application. On the upper left, the camera input image is located. On the lower left, the matched reference image is displayed. On the right hand side, the browser window can be seen. There, the description of the object, associated to the matched model image, is shown.

of the illuminant. Furthermore, the difference of quality between the reference images in the database and the images taken with the low-quality webcam affect the results only to a limited extent. However, input images with low contrasts are difficult to recognise and mainly these lead therefore often to mis-recognitions.

Note that in contrast to the approach described in [5], we do not use colour information for the object recognition. This is one of the reasons for the above-mentioned recognition robustness under various lighting conditions. We experimentally verified that illumination variations, caused by artificial and natural lighting, lead to low recognition results when colour was used as additional information.

5. CONCLUSION

In this paper we have described the functionality of an interactive museum guide. It allows to robustly recognise museum exhibits under difficult environmental conditions. Furthermore, our guide is robust to changes of the viewing angle as well as rotation and scale. The museum guide is running on a standard low-cost hardware. Moreover, we presented a possibility to improve the accuracy and speed of object recognition by combining image-based object recognition with automatic room detection.

Future work will be focused on developing an even faster algorithm for the matching task. Furthermore, we want to deploy hardware that is better suited for the task at hand. Specifically, we will test cameras which provide images of higher quality and we also attempt to implement our approach onto a smaller portable device.

6. ACKNOWLEDGEMENTS

The authors acknowledge support by the ETH project of versatile computing, the EC Network of Excellence EPOCH and the Swiss NCCR project IM2. We also gratefully acknowledge the crucial support by the Swiss National Museum Zurich.

7. REFERENCES

[1] <http://www.btnode.ethz.ch>.

- [2] A. Baumberg. Reliable feature matching across widely separated views. In *Computer Vision and Pattern Recognition*, pages 774–781, 2000.
- [3] J. Beis and D. G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.
- [4] W. Burgard, A. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. The interactive museum tour-guide robot. In *Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, 1998.
- [5] P. Föckler, T. Zeidler, and O. Bimber. Phoneguide: Museum guidance supported by on-device object recognition on mobile phones. Research Report 54.74 54.72, Bauhaus-University Weimar, Media Faculty, Dept. Augmented Reality, 2005.
- [6] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition*, pages 506–513, 2004.
- [7] F. Kusunoki, M. Sugimoto, and H. Hashizume. Toward an interactive museum guide with sensing and wireless network technologies. In *WMTE2002, Varjo, Sweden*, pages 99–102, 2002.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints, cascade filtering approach. *International Journal of Computer Vision*, 60(2):91–110, January 2004.
- [9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Computer Vision and Pattern Recognition*, volume 2, pages 257–263, June 2003.
- [10] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *International Journal of Robotics Research*, 19(11):972–999, 2000.
- [11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, 2001.