Mo Músaem Fíorúil: A Web-based Search and Information Service for Museum Visitors

Michael Blighe, Sorin Sav, Hyowon Lee and Noel E. O'Connor

Centre for Digital Video Processing, Adaptive Information Cluster, Dublin City University, Ireland
blighem@eeng.dcu.ie
WWW home page: http://www.cdvp.dcu.ie

Abstract. We describe the prototype of an interactive, web-based, museum artifact search and information service. Mo Músaem Fíorúil clusters and indexes images of museum artifacts taken by visitors to the museum where the images are captured using a passive capture device such as Microsoft's SenseCam [1]. The system also matches clustered artifacts to images of the same artifact from the museums official photo collection and allows the user to view images of the same artifact taken by other visitors to the museum. This matching process potentially allows the system to provide more detailed information about a particular artifact to the user based on their inferred preferences, thereby greatly enhancing the user's overall museum experience. In this work, we introduce the system and describe, in broad terms, it's overall functionality and use. Using different image sets of artificial museum objects, we also describe experiments and results carried out in relation to the artifact matching component of the system.

1 Introduction

The traditional museum visitor experience has been characterized by having to choose between a limited number of predefined guided tours and the challenge of visiting on one's own. Despite the stimulating environment created in museums, they often fall short of supporting their visitors, either before, during, or after the visit, in terms of analyzing and learning about what's been seen and found to be of interest. One way museums have attempted to tackle this problem is with the increasing use of audio guides. Visitors select audio sequences related to particular pieces by keying in a code associated with a particular artifact, or perhaps a particular exhibition space within the museum. In the latter case, the visitor is then guided around the space by the audio guide, visiting the artifacts in a pre-planned way, designed by museum personnel. One obvious advantage of audio guides is their availability in multiple languages. However, the use of audio guides presents several drawbacks for both the museum and its visitors. The audio guide can present auditory information only, excluding other forms of communication which may be more interesting, or more effective, such as written text, images, video, or interactive applications. They require the visitors to key in the code for each exhibit and this can become frustrating and detract from the overall experience.

One possibility for making exhibitions more attractive to the visitor is to improve the interaction between the visitor and the objects of interest by means of supplementary information either during or after the visit. Targeting the latter initially, we have developed an interactive museum information Web 2.0 prototype system that is able to automatically index and retrieve information about the objects a visitor found interesting. The visitor, wearing a passive image capture device, generates images of various artifacts whilst wandering around the museum. The device could be supplied by the museum and retrieved at the end of each visit, thereby ensuring control over the image collections generated. In order to subsequently access their personalised museum tour via the museum's web site, visitors need only be supplied with a unique username. Once they get home, they can log on to the museum website and relive their museum experience by browsing their photos and automatically recommended supplementary material, chosen based on their known interactions. Given that the system can determine which particular artifacts the user visited, additional information (e.g. sketches, 3D models, explanatory text, professional photos, etc) about a particular object could be provided to the user, as well as images other visitors have captured of the same artifact. This has the ancillary benefit of increasing usage of museum web-resources and providing web access to museum catalogues, but not at the expense of deterring visitors – a key concern for museums when considering web-based services.

The choice of passive image capture as a means of gathering data about a visitor's museum experience carries with it two key advantages. Firstly, it provides a visual record of what the user *saw* as opposed to simply where he/she *was*. This potentially allows us to infer user preferences and interests in a more finely grained manner than location-based sensing (e.g. we could determine which specific artifact in a display case was of particular interest). Secondly, adopting such capture devices is a relatively straightforward extension of audio tours that does not require any re-engineering of the museum infrastructure itself unlike, say, a costly installation of a RF-ID tracking system. Of course, this technology choice also brings its own challenges. Not least of these is the uneasiness (or in fact outright refusal in many cases) of many museums to allow image capture of their collections. However, this is changing slowly as museums consider new business models based on emerging technology and indeed some museums have already embraced the concept of personal image capture for non-professional/commercial purposes. London's Tate Modern has recognised the advantages of encouraging it's visitor's to interact (email content home, respond to questions, etc.) with the exhibits [2]. The Rijksmuseum in Amsterdam is exploring similar possibilities [3]. One goal of our work is to demonstrate the possibilities of state-of-the-art imaging technology in this application context. In this paper, we briefly describe the overall functionality of the proposed system, however, our main focus is on the artifact recognition component and the initial proof of concept results obtained with a small number of artificial museum artifacts.

The rest of this paper is organized as follows. In Section 2, we review related work in this area. We introduce our system in Section 3 and discuss the artifact matching system used in this work in Section 4. In Section 5, we outline the experiments performed and results obtained. This is followed by a discussion in Section 6, whilst future work and conclusions are discussed in Section 7.

2 Related Work

The value of multimedia for a mobile museum guide is discussed by Proctor & Tellis [2] who present an extended user study conducted at the Tate Modern in 2002. They highlight the components necessary for a successful museum installation: content, user interface, applications, form factor and positioning. Fockler et al. developed PhoneGuide [4], a system which supports on-device object recognition on a mobile phone. They extracted a number of low-level colour features and classified the images using a single layer perceptron neural network. However, the use of colour features means that their system is not robust to changes in lighting, viewpoint and illumination. In addition, the method is not compatible across different models of mobile phone due to the varying colour responses of the cameras used in different models.

Bay et al. [5] proposed an Interactive Museum Guide using a tablet PC with a touchscreen, a webcam, and a bluetooth receiver. The guide recognizes objects on display in the museum based on images of particular artifacts taken directly by the visitor. The system then displays additional information to the user about the object in question. In addition, the system can determine the visitor's location by receiving signals emitted from bluetooth nodes located in different display areas throughout the museum. This information is used to reduce the search space for the extraction of relevant objects. A similar system using infrared for location identification was developed by [6]. However, these systems all require a certain level of infrastructure to be built and maintained within the museum and many museums may be reluctant to accept these technological intrusions. In addition, the current hardware platforms used are cumbersome and are not practical in real scenarios. Other approaches include robots that guide users through museums [7]. However, these are not appropriate for individual use and are difficult to adapt to different environments.

Each of the described approaches involves the introduction of a novel piece of hardware equipment into the museum environment (i.e. robots, PDA's, mobile phones, tablet PC's). The presence of the technology itself changes the dynamic of the museum experience for the visitor as described by Semper et al. [8]. They describe how the introduction of handheld devices to the museum environment tended to distract people from freely using their hands to interact with the exhibits (a highly interactive science museum in their study). This is due to the fact that the visitor had to actually hold the device in their hands, and some were afraid of damaging an expensive piece of equipment [9], and the visitor spent more time reading the content displayed on the device, as opposed to focusing on the information and artifact's on display in the museum. They also found that using handhelds disrupted the normal social interactions between members of social groups. This is analogous to similar social consequences arising from the use of mobile phones in every day life. We believe that these issues can be overcome by using passive capture devices to record the user's visit and experience.

Passive capture devices are cameras which automatically take pictures without any user intervention [10] [11]. They are ideal for use in a museum environment as they allow the visitor to record their experiences without conscious thought. The advantages of this method of capturing photos are increased coverage of, and improved participation in, the event itself. However, the passive capture of photos presents new problems, particularly, how to manage and organise the massively increased volume of images captured [12]. Traditional systems for content-based image retrieval are not up to this task. In [11] the authors describe the MyLifeBits system, which is a first step in tackling this problem, specifically in relation to the images captured by SenseCam. MyLifeBits also captures other forms of digital media and is a step toward's fulfilling Bush's Memex vision [13]. Other forms of passive capture devices include the StartleCam [10] and the Campaignr project [14]. Campaignr is a software framework for mobile phones that enables owners of smartphones (specifically Symbian Series 60 3rd edition phones) to participate in data gathering campaigns. We use images captured by the SenseCam and a Nokia N95 running the Campaignr software in this work. Both the SenseCam and N95 are worn around the visitors neck to allow the capture of images in a passive manner (see Figure 1).

Fig. 1. User shown wearing the SenseCam

The task of identifying similar artifacts within a database of images remains challenging due to viewpoint or lighting changes, deformations, and partial occlusions that may exist across different examples. Global image features, based on image properties such as colour or texture, have proven to be of limited use in these real-world environments. Instead, researchers have recently turned to representations based on local features that can be reliably detected and are invariant to the transformations likely to occur across images (i.e. photometric or various geometric transformations).

One approach has been to use a corner detector to identify repeatable image locations, around which local image properties can be measured. Schmid et al.

[15] developed one of the earliest object matching systems using these features. They extracted local gray value feature points with a Harris corner detector, and then created a local image descriptor at each interest point. These image descriptors were used for robust object recognition by looking for multiple matching descriptors that satisfied object-based orientation and location constraints. However, this approach only examined an image at a single scale. As the change in scale becomes significant, these detectors respond to different image points.

More recently, there has been great progress in the use of invariant features [16] [17] for object matching. With these features, robustness to small changes in viewpoint as well as to partial occlusion is achievable and objects can be recognized anywhere in an image, with arbitrary size, rotation, and without using a previous object segmentation step [18]. It follows, therefore, that these features can be matched more reliably than traditional methods such as cross-correlation using Harris corners.

3 Museum Information System

Mo Músaem Fíorúil (My Virtual Museum in the Irish language) is a web-based museum artifact search service where the users of the service, after visiting a museum and taking a number of photos of artifacts, can upload their photos to a website and find information about the artifacts those photos had captured. On its web interface, a user's uploaded photos are displayed with the groupings of photos automatically formed based on the unique artifacts among the photos, and the user can drag and drop the photos into different groupings if wished. Once a particular grouping that features a unique museum artifact is selected, the system presents a list of museum artifacts that matches the user's photos, and selecting one of these will present full information about the artifact. Another way to view the interaction paradigm of this service is that the museum visitor can use their photos as query images to the service, and the retrieval result shows full information about the artifacts those photos contain.

Two passive capture devices were used to acquire the images used in this system - the Microsoft SenseCam and a Nokia N95 running the Campaignr software. Should users wish to manually capture an image, they can do so using the SenseCam, by simply pressing a button on the side of the camera, or by using the N95 in the traditional manner in which camera phones operate. In this initial prototype, artificial artifacts have been used with images captured in a lab environment. The artifacts are limited in size to $30 \times 20 \times 30$ cm, due to the constraints imposed by our object model capture system (see section 4). The descriptions of the recognised artifacts are fictional and are intended to simulate the workings of a real system. Once the user has selected an artifact of interest, the system will also show the pre-captured model of the artifact, that the user can rotate 360°. Images that other users have taken of the same object and which may also be of interest are also displayed. The system is freely accessible online for demonstration purposes (http://www.eeng.dcu.ie/˜vmpg/ksDemo/ks.html).

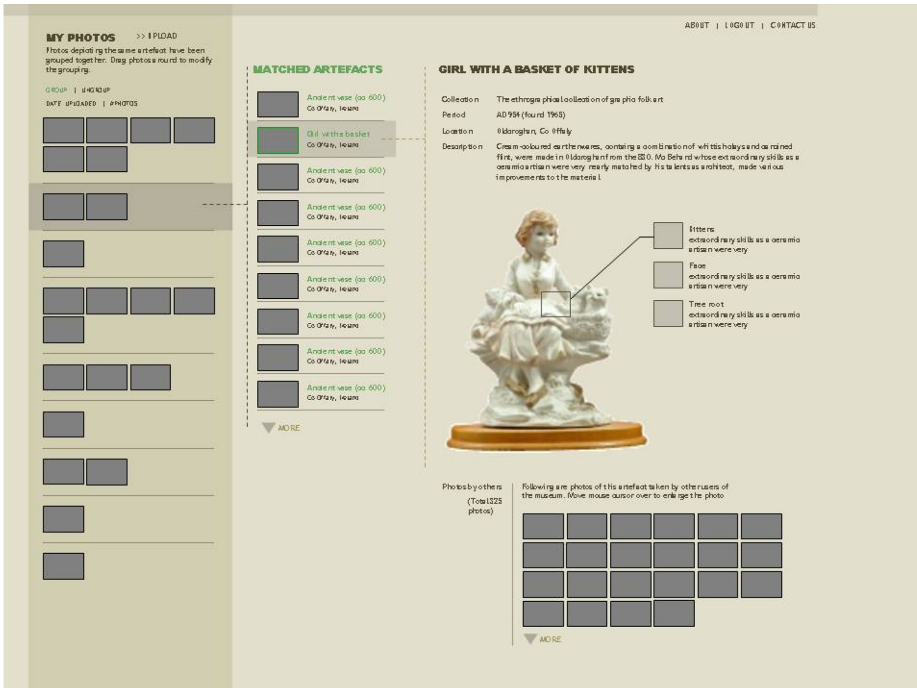

Fig. 2. Museum Information System

In order to demonstrate the artifact matching capabilities of our system, we created a database with artifical museum objects. The database contains images of 10 different objects, taken from multiple viewpoints with lighting, rotation and scale changes. A sample image of each of the 10 chosen objects is shown in Figure 3.

4 Object Matching System

Model images are generated using a static camera rig and an automated turntable. The turntable is situated in a light tent with diffuse ambient lighting and a controlled known-colour background. Each artifact is placed on the table and captured as it is rotated. The object is then segmented from the background using a straightforward chroma-keying process.

In order to perform matching, we utilize an approach similar to that outlined by [17]. This approach uses the SIFT local descriptors as they have proved well-adapted to matching and recognition tasks as they are robust to partial visibility and clutter. Mikolajczyk et al. [19] have compared several descriptors for matching and found that SIFT descriptors perform best so we continue with SIFT on this basis. In order to perform object matching, we follow the following procedure. First, the SIFT features are computed from the input image. Each

Fig. 3. Sample images of the 10 artificial artifacts

keypoint is then independently matched to the database of keypoints extracted from the training images. This feature matching is done through a Euclidean-distance based nearest neighbor approach. Many of these initial matches will be incorrect due to ambiguous features or features that arise from background clutter. To increase robustness, matches are rejected for those keypoints for which the ratio of the nearest neighbor distance to the second nearest neighbor distance is greater than 0.8. This discards many of the false matches arising from background clutter. Finally, to avoid the expensive search required for finding the Euclidean distance based nearest neighbor, an approximate algorithm, called the Best-Bin-First (BBF) algorithm [20] is used. This is a fast method for returning the nearest neighbor with high probability. For a database of 100,000 keypoints, this provides a speedup over exact nearest neighbour search by about 2 orders of magnitude yet results in less than a 5% loss in the number of correct matches.

Although the distance ratio test described above discards many of the false matches arising from background clutter, we can still have matches that belong to different objects. Therefore to increase robustness to object identification, we want to cluster those features that belong to the same object and reject the matches that are left out in the clustering process. This is done using the Hough Transform [21]. Each keypoint specifies 4 parameters: 2D location, scale and orientation. Using these parameters we use the Hough Transform to identify clusters of features that vote for the same object pose. The probability of the interpretation being correct is much higher than for any single feature. Each keypoint votes for the set of object poses that are consistent with the keypoint's location, scale, and orientation. Bins that accumulate at least 3 votes are identified as candidate object/pose matches. Therefore, clusters of at least 3 features are first identified that agree on an object and its pose, as these clusters have a much higher probability of being correct than individual feature matches. Then, each cluster is checked by performing a detailed geometric fit to the model, and the result is used to accept or reject the interpretation.

For each candidate cluster, a least-squares solution for the best estimated affine projection parameters relating the training image to the input image is obtained. If the projection of a keypoint through these parameters lies within half the error range that was used for the parameters in the Hough transform bins, the keypoint match is kept. If fewer than 3 points remain after discarding outliers for a bin, then the object match is rejected. The least-squares fitting is repeated until no more rejections take place.

5 Experimental Results

A number of experiments were carried out on different combinations of test and model images. We created 3 sets of model images. The reasons for the choice of three different model sets were the use of two different cameras and in order to determine if the effort required to segment the artifacts from the background using the static camera rig (see section 4) was justified. The first set of model images, labeled *m1*, were captured using the static camera rig. This created images of size 320×240, taken from 12 different viewing angles around the artifact, of each of the 10 artifacts in our database. This allows for a greater degree of view-point independence. Due to the fact that our training images were all taken from different viewing angles, we only use 5 of these images in this model set (although the 12 images are used to *rotate* the artifact on the user interface) (see Figure 4). This gave a total of 50 model images.

Fig. 4. Example of the 5 model images for one of the 10 artifacts

The second set of model images, labeled *m2*, contained 3 SenseCam images for each artifact in the database, taken from 3 different viewing angles in front of the artifact in question. This gave a total of 30 model images. The final model collection, *m3*, consisted of 10 images (1 for each artifact) taken with the higher resolution Nokia N95 camera. Sample images from *m2* and *m3* can be see in Figure 5.

We used two different test sets, one for each of the cameras used. 100 images of size 640×480 were taken with the Microsoft SenseCam and 100 images of size 2592×1944 with the Nokia N95. Each set contains multiple images of all objects with differing scale, rotation, viewpoint and lighting conditions. Images were captured by simulating a museum visitors inspection of the artifacts. The objects used are made of different materials, have different shapes, and include ceramic vases, statues and jugs, metal and stone items, and a teddy bear. Some of the objects were placed on a glass table which produced interfering reflections.

Fig. 5. Example of SenseCam (1^{st} row) & N95 (2^{nd} row) model images

Each test image set was evaluated on each model set, giving a total of 6 different sets of experimental results. We used the confusion matrix in order to evaluate our results, however, due to constraints on space we only show 4 sets of results (shown in Tables 1-4). The results obtained using the model set *m1* were omitted as they were the poorest. The remaining results represent those obtained against the second, *m2*, and third, *m3*, sets of model images. The significant difference between these two sets of model images and set *m1* is that the background is available and, therefore, provides features for recognition.

The results varied considerably across each combination of test and model sets of images. The *Footballer* proved challenging across all experiments. The highest recognition rate achieved for this artifact was only 40% using SenseCam test images and the *m2* set of model images. Other objects, such as the *Statue*, could not be detected at all using SenseCam and the *m1* set of model images, but achieved recognition rates of 80% using SenseCam test images and model images *m2*. Recognition rates of 100% were obtained for the *Striped Vase* and *Vinegar* using N95 test images and the *m3* set of model images. In general terms, the worst performing results were those obtained using the set of images captured using the static camera rig (*m1*) for both cameras. The best sets of results were obtained when both the test and model images were taken with the same cameras. However, impressive results can also be seen using test and model images from different cameras.

True classes	**Teddy**	**Cellar**	**Floral vase**	**Blue jug**	**Footballer**	**Navy Jug**	**Plaque**	**White Statue**	**Striped vase**	**Vinegar**
Teddy	**80**	0	0	0	0	0	20	0	0	0
Cellar	20	**60**	0	0	0	0	20	0	0	0
Floral vase	30	0	**40**	10	0	0	10	0	0	10
Blue jug	30	0	0	**50**	0	0	10	10	0	0
Footballer	60	0	0	0	**40**	0	0	0	0	0
Navy jug	10	0	0	20	0	**50**	20	0	0	0
Plaque	0	0	0	0	0	0	**90**	0	10	0
White statue	0	0	0	10	0	0	10	**80**	0	0
Striped vase	20	0	0	0	10	0	10	0	**60**	0
Vinegar	40	0	0	0	0	0	0	0	40	**92**

Table 1. Confusion matrix for SenseCam test and model images

True classes	Teddy	Cellar	Floral vase	Blue jug	Footballer	Navy Jug	Plaque	White Statue	Striped vase	Vinegar
Teddy	**80**	0	10	0	0	0	10	0	0	0
Cellar	20	**50**	10	0	0	0	0	10	20	0
Floral vase	40	0	**10**	0	0	0	0	10	30	10
Blue jug	20	0	0	**30**	0	0	0	10	20	20
Footballer	0	20	10	0	**10**	0	20	10	30	0
Navy jug	11	11	0	11	0	**23**	0	11	33	0
Plaque	10	0	0	0	0	0	**90**	0	0	0
White statue	20	20	0	0	0	0	0	**40**	20	0
Striped vase	10	0	10	0	0	10	0	0	**70**	0
Vinegar	0	8	8	0	0	0	0	0	40	**84**

Table 2. Confusion Matrix for N95 test and SenseCam model images

True classes	Teddy	Cellar	Floral vase	Blue jug	Footballer	Navy Jug	Plaque	White Statue	Striped vase	Vinegar
Teddy	**80**	0	0	0	0	0	20	0	0	0
Cellar	0	**70**	0	0	0	0	20	0	0	10
Floral vase	20	10	**50**	0	0	0	20	0	0	0
Blue jug	10	10	10	**40**	10	0	10	0	0	10
Footballer	20	10	10	0	**30**	0	20	0	0	10
Navy jug	11	0	0	0	11	**78**	0	0	0	0
Plaque	10	0	0	0	10	0	**80**	0	0	0
White statue	60	10	10	0	0	0	10	**10**	0	0
Striped vase	0	0	0	0	0	0	0	0	**100**	0
Vinegar	0	0	0	5	0	0	0	0	0	**100**

Table 3. Confusion matrix for N95 test model images

6 Discussion

The poor results obtained using the segmented model images (*m1*) was surprising, as this is an approach often taken in the object recognition literature. However, in many of the test images the artifacts were extremely small in size meaning that the image contained a lot of background. In many of these cases, the algorithm found more matches on the background objects leading to a matching failure. These initial results would therefore suggest that the effort required to remove the background from the images, using the static camera rig, is not justified.

The importance of including the background as part of the model image can be seen in the improvement in results using the remaining sets of model images. Certain artifacts were successfully matched despite variations in lighting, scale, rotation and viewpoint. However, the recognition performance for others was quite low. This was again due to the background, however, it was caused by deficiencies in our experimental setup. Certain artifacts were taken in exactly the same location (i.e. we placed one object on the surface, captured images of it, and then replaced it with the next artifact). This meant that the background information in certain groups of artifacts was the same. In situations where the artifact did not provide enough robust or discriminant features, the background

True classes	Teddy	Cellar	Floral vase	Blue jug	Footballer	Navy Jug	Plaque	White Statue	Striped vase	Vinegar
Teddy	**60**	0	0	10	10	0	10	10	0	0
Cellar	30	**60**	0	0	0	0	20	0	10	0
Floral vase	20	10	**0**	20	10	10	20	0	10	0
Blue jug	10	0	20	**30**	0	0	20	0	10	0
Footballer	10	20	0	0	**30**	20	0	0	0	10
Navy jug	0	10	0	0	0	**90**	0	0	0	0
Plaque	10	10	0	0	0	0	**70**	0	10	0
White statue	30	10	0	0	0	0	20	**30**	0	10
Striped vase	0	0	10	0	0	0	0	0	**90**	0
Vinegar	10	10	20	0	0	0	10	0	20	**30**

Table 4. Confusion matrix for SenseCam test and N95 model images

information was used to match the image. In many cases, the background was matched to the same background object but the image contained a different artifact captured in the same location. In particular, if we examine table 1 we can see that many artifacts have been incorrectly classifed as being an instance of a *Teddy* or a *Plaque*. This was due to this particular issue and further testing should yield improved results in this regard. In a realistic museum setting, this problem should not occur.

7 Conclusions

We have presented a novel system for providing visitors to a museum a means of interacting with and learning more about their visit. The system incorporates a passive capture camera device and a web-based user interface. The camera captures images continuously during the museum visit. These can then be uploaded to the system, via the website, allowing the user to browse their own photo collection, match their images to images from the museums private image collection, access more detailed information concerning artifacts of interest, and also view images other visitors to the museum have taken of similar artifacts. We also described in detail the operation of the artifact matching element of the system and presented some experimental results. This element of the system uses SIFT images features which are robust to changes in lighting, scale and rotation.

Much future work remains. As we extend to more museum artifacts, the matching accuracy and speed of the system will decrease as many more similar artifacts are added. In addition, more background clutter could lead to more false detections. We plan to explore the use of location based methods in order to assist us in reducing the search space necessary to match in a database of many more musuem artifacts.

Acknowledgements

The research leading to this paper was supported by the European Commission under contract FP6-027026 (K-Space), Microsoft Research and Science Foundation Ireland under grant number 03/IN.3/I361. We would like to express our gratitude to CENS at UCLA for providing us with the Campaigr software and support, and the personnel of the National Museum of Ireland for providing valuable insights to assist us in the development of all aspects of the proposed system.

References

1. Gemmell, J., Williams, L., Wood, K., Lueder, R., Bell, G.: Passive capture and ensuing issues for a personal lifetime store. (October 2004)
2. Proctor, N., Tellis, C.: The state of the art in museum handhelds in 2003. In: Museums and the Web. (2003)

3. Aroyo, L., Wang, Y., Brussee, R., Gorgels, P., Rutledge, L., Stash, N.: Personalized museum experience - the rijksmuseum use case. In: Museums and The Web. (April 2007)
4. Fockler, P., Zeidler, T., Brombach, B., Bruns, E., Bimber, O.: Phoneguide - museum guidance supported by on-device object recognition on mobile phones. 54 74, Bauhaus-University Weimar, Weimar, Germany (2005)
5. Bay, H., Fasel, B., Gool, L.V.: Interactive museum guide - fast and robust recognition of museum objects. In: First International Workshop on Mobile Vision. (2006)
6. Oppermann, R., Sprecht, M.: Adaptive mobile museum guide for information and learning on demand. In: 8th International Conference on Human-Computer Interaction. Volume 2. (1999) 642–646
7. Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: Probabilistic algorithms and the interactive museum tour-guide robot minerva. International Journal of Robotics Research **19**(11) (2000) 972–999
8. Semper, R., Spasojevic, M.: The electronic guidebook - using portable devices and a wireless web-based network to extend the museum experience. In: Museums and the Web. (April 2002)
9. Hsi, S.: The electronic guidebook - a study of user experiences using mobile web content in a museum setting. In: International Workshop on Wireless and Mobile Technologies in Education. (August 2002)
10. Healey, J., Picard, R.: Startlecam: A cyberbetic wearable camera. (October 1998)
11. Gemmell, J., Aris, A., Lueder, R.: Telling stories with mylifebits. ICME (July 2005)
12. Blighe, M., Borgne, H.L., O'Connor, N.: Exploiting context information to aid landmark detection in sensecam images. In: 2nd International Workshop on Exploiting Context Histories in Smart Environments - Infrastructures and Design (ECHISE). (September 2006)
13. Bush, V.: As we may think. The Atlantic Monthly (July 1945)
14. Joki, A., Burke, J., Estrin, D.: Campaignr - a framework for participatory data collection on mobile phones. Technical Report 770, Centre for Embedded Network Sensing, University of California, Los Angeles (October 2007)
15. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(5) (1997) 530–535
16. Baumberg, A.: Reliable feature matching across widely separated views. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. (2000) 774–781
17. Lowe, D.: Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision. Volume 60(2). (2004) 91–110
18. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. International Journal of Computer Vision **37**(2) (2000) 151–172
19. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(10) (February 2005) 1615–1630
20. Beis, J., Lowe, D.: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Conference on Computer Vision and Pattern Recognition. (1997) 1000–1006
21. Ballard, D.: Generalizing the hough transform to detect arbitrary patterns. Pattern Recognition **13**(2) (1981) 111–122