

What is a feasible method to transform
contextualized textual architectural concepts
into visual markers to be presented in real
time as overlays in AR environments?

O.R. Jozefzoon

Department of Artificial Intelligence
Faculty of Physics, Mathematics and Computer Science
Amsterdam, North-Holland, The Netherlands

June 15, 2013

Abstract

1 Introduction

While much research has been conducted on 2D object retrieval, an increased interest for 3D object retrieval is also observable.

The Museum of Amsterdam, for example, has created the Museum App¹ for mobile phones, which currently operates within a 2D environment.

To improve the user experience a decision has been made to incorporate Augmented Reality into the device to transform the existing data from the tour to a live, direct or indirect, view of a physical, real-world (3D) environment of which the elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS data. It is within this new

¹An enhanced museum guidance system that uses trigger points, information nodes related to GEO points, to enrich the user with content during a tour around the city of Amsterdam.

environment that the user will be able to find a given object, as described in the tour, by means of the method of highlighting recognized patterns, in real-time.

This research focuses on finding the most suitable manner for adding the Augmented Reality feature to the existing application. However, due to time constraints, the main focus lies on finding one particular object, in this case 'coat of arms', abstracted from the textual tour data. For reasons that will be explained throughout the rest of this paper, this research uses HOG features (Dalal et al. 2005) to detect an object and SIFT (Lowe, 1999) to recognize an object in locale architecture regardless the angle between the observer and the object. Even though we use HOG and SIFT as depictees because they are widely used to retrieve 2D objects, problems occur when opting for the recognition of 3D objects. The detection of one of these problems will be presented at the end of the paper and is believed to contribute to the cause for implementing effective 3D object retrieval systems.

Outline The remainder of this article is organized as follows. Section ?? gives account of previous work to determine what is already available as potential solutions to this research. After which a description of the missing aspects on which the research question is based is presented in section 3. Section 4 then outlines the method and approach used to answer the research question and presents the added feature to the existing tour guidance application. Finally section 5 covers the evaluations, presents the results and report the finding of this research.

2 Related research

Many applicants of location-aware applications propose the use of trigger zones to activate content and the use of servers to perform object recognition. Even though, location-awareness by the use of trigger zones is a highly efficient manner to display or use content, this dependency does not allow much room flexibility of the application. Lameira et al. (2011) [?] propose an application for real time object recognition without the use of a server. Even though this paper does not completely ignore the value of a server, the aim is to limit the use of the server to a maximum and to rely on the computations power of the mobile device. This decision is based on the fact that mobile device, nowadays, are improved, have increased computational power

and are, thus, able to perform complex computations within a reasonable amount of time. This approach limits the use of a server to a storage facility for the data and allows for the application to perform as independent and as fast as possible².

Amlacher et al. (2008) [?] demonstrate in their paper geo-indexed object recognition from experimental tracks and image captures in an urban scenario, extracting object hypotheses in the local context from both (i) mobile image based appearance and (ii) GPS based positioning. This paper also proposes to use of geo-indexed verification to determine whether or not the assumptions, based on extracted object hypotheses conceptualized from text, made by our application are accurate. The results from the demonstration by Amlacher et al. (2008) provide us with a basis assumption to use geo-indexed data to improve the accuracy and coverage of our application. An accuracy of 92% in recognizing object in general is desirable. By using geo-indexed verification the application will be able to restore images, which contain parts of the Coat of Arms, from memory and combine them into a complete image of a Coat of Arms. This approach will provide the application with an extra device to achieve the highest possible accuracy for the detection and recognition of Coat of Arms.

The client-side tracker as proposed by Gammeter et al. (2010) [?] in Server-side object recognition and client-side object tracking for mobile augmented reality is a valuable device, which we will use to memorize the position of an object even when out of screen, using visual and sensor based cues. This client-side tracker is necessary because the user will be guided towards an object to be recognized. Gammeter et al. do not use GPS information for object recognition and tracking. This paper, on the other hand, will use GPS information for the verification of the results presented by the application and, combined with the compass on the mobile device, as a device to guide the user in to the desired direction. The use of the GPS and Compass information will provide the application with a more robust base to guide the user towards the object and to co-operate with the object recognition algorithm to achieve the highest accuracy in detecting and recognizing of a Coat of Arms.

Finally, as described in the introduction, this paper proposes the use of

²Independent in the sense that the application is not dependent of an internet-connection to connect with the server, once the data is stored, and as fast as possible in the sense that the computation does not depend on an external source, but is almost completely self-sufficient.

HOG features (Dalal et al. 2005) combined with SIFT (Lowe, 1999) to detect and recognize Coat of Arms in locale architecture. The combination of these algorithms is based on the fact that SIFT uses local contrasts to differentiate between features of selected points. This means that for our images we will achieve poor results in detection, because the images in our database do not contain sufficient local contrasts. However, with the use of HOG features it is possible to overcome this weakness and to collect a collection of unique points based on their edge orientation.

The description of the architectural concepts needs some form of accepted base. For this task the Getty's AAT³ can be used. AAT stands an acronym for Art and Architectural Thesaurus and is a

controlled vocabulary used for describing items of art, architecture, and material culture.
... structured vocabulary of around 34,000 concepts, including 131,000 terms, descriptions, bibliographic citations, and other information relating to fine art, architecture, decorative arts, archival materials, and material culture.⁴

3 Research question.

Based on the presented literature and the established problem of the Amsterdam Museum we decided on the following research question:

What is a feasible method to transform contextualized textual architectural concepts into visual markers to be presented in real time as overlays in AR environments?

Our aim is to reduce the dependency of the triggers by implementing an application that provides a user with content based on an observation of an object in the real world (i.e. through a camera of the mobile phone) and on abstraction from a textual source (the concepts highlighted by the user in the Amsterdam Museums Tour). The transformation process from textual concept into the final marker will adapt AAT descriptions of object/buildings, in combination with GEO data and compass information. In that way we propose a new approach in which our content is activated by the input from

³<http://www.getty.edu/research/tools/vocabularies/aat/>

⁴Source: <http://en.wikipedia.org/wiki/AAT>

the data (text/image) and afterwards verified using geo-location and compass data from the mobile device.

4 Method and approach

In order to achieve our goal, of creating a feasible model that transforms contextualized textual architectural concepts into visual markers to be presented in real time as overlays in AR environments, we will use the method of implementing and testing. In the first stage we will decide for a manner to conceptualize the textual data to a model which will be useful for the transformation process. Here the conceptualization will be determined with the use of descriptions from the AAT and manually selected locations. The locations will serve as a reference- or pinpoints on which we will base our conceptualization. This conceptualization will be necessary, because our algorithm has to be provided with a base to work on. Once the correct representation of our data is determined we will start with the implementation of our application. The application will use the conceptualized data to present the user with the correct or clear description, as shown in figure 1, of an object in the real world. In other words, the conceptualized textual data will be transformed into a usable model that can be used to recognize certain patterns in videoimages and highlights the recognized patterns in the screen of the user.

When it seems as if the correct description is not provided, thus none or the wrong object are highlighted, we will explore the added feature of using the GPS and/or compass data for an extra verification. This verification might be necessary when the user, for example, does not find the object described or the application does not find a matching description for the object presented by the user.

5 Discussion

As mentioned beforehand, in this section we will describe how will the results of the research be evaluated? In a way this can be regarded as part of the method/approach, but it is important and therefore requires independent attention.

The main evaluation will occur after finishing the implementation of the



Figure 1:
Image obtained from http://www.ellipsead.com/mobile_ar to demonstrate the eventual output, the highlighting of objects in an image.

application. However, we will incorporate unit test in our code to make sure that we are not providing our application with incorrect data our malicious chunks of codes. The unit tests are necessary for our application, because malicious codes may corrupt the accuracy and might negatively affect the coverage. For our main evaluation real world testing will be necessary. This means that we will test if the application works as desired on the selected locations. If we are not provided with the desired results then we will have to decide whether the source of the error is our conceptualization of the text or whether the error could be found at the side of the object recognition algorithm(s) used. In case of the first we have to reconsider our conceptualization, in case of the latter an obvious reconsidering of algorithm will be necessary. We will probably mainly revise the algorithms, because the conceptualization will be given a fair amount of (re-) consideration and once the decision is made for an acceptable representation it is our task to implement a suiting algorithm.