

Augmented Reality for Tourism in Oman Using Free Open Source Software

Ahmed Al Maashri, Sara Al-Asadi, Muna Tageldin, Shurooq Al-Lawati, Ali Al Shidhani

Abstract— Tourism is an important and key economic driver in many parts of the world due to its high revenues. Tourism in Oman is induced by its spectacular beauty thanks to the geographic and environmental diversities that it contains. Therefore, Oman is looking forward to increase the contribution of tourism sector to the country's Gross Domestic Product. Because of these reasons, there are needs for creative tourism services to keep the flow of tourists coming to the country. Due to the diversity and disparity of tourist landmarks, there is a need for a portable and smart information dissemination service to provide rich information about the landmarks. The service has to be easily accessible, timely, and accurate. This paper presents the Portable Tour Guide, a free and open source solution targeting the tourism sector in Oman. The guide utilizes Augmented Reality technology to enhance the experience of the tourist using cloud computing services. Results of experiments on the implemented system show good accuracy and short response time.

Index Terms— Augmented Reality, Oman Tourism, FOSS

XXV. INTRODUCTION

ACCORDING to *Oman Observer* newspaper, more than two (2) million tourists from different nations have visited Oman in 2012. Compared to 2011, there is a significant increase by 48% [1]. Due to the importance of the tourism sector in Oman's economic diversification, it is forecasted that by the year 2024, tourism sector will contribute by 3.9% to the country's total Gross Domestic Product (GDP) [2].

However, one of the challenges that tourists face during their visit is the lack of tour guidance and proper information dissemination. For example, tourists may be interested in learning about the historical background of the landmark they just visited. Unless a professional tour guide is hired, it is difficult for the tourists to learn about the landmark and appreciate its significance to the country and to the world's heritage.

Therefore, a technology-compatible system is needed to provide the required information associated with landmarks. The objective of such a system is to enhance the tourists' experience by making the landmarks more appealing. This system should provide the tourists with a visual landmark-identification facility during their visit to Oman in real-time. More importantly, Augmented Reality (AR) technology is utilized to enhance the experience of the tourists by

superimposing useful information on the image of the landmark when viewed on the tourists smart phone.

We present Fig. 1 to demonstrate how AR is envisioned to guide the tourists in their visits. The figure shows a display screen of a smart phone, where the tourist has taken a picture of the landmark; in this case "Zawawi Mosque" in Muscat, Oman. Then, using a cloud computing service that is dedicated to this purpose, the smart phone will receive useful information that is superimposed on the original picture. As shown in the figure, a brief description is provided to the tourist.

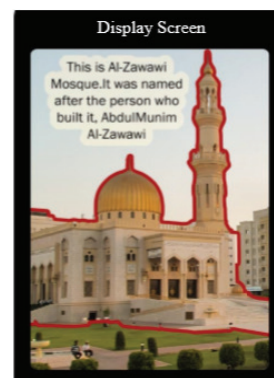


Fig. 1 A demonstration of the proposed system that uses AR technology to augment landmarks with useful information

The remainder of this paper is organized as follows: Section II presents related work, followed by a discussion of embedded vision systems and AR technology in Section III. Section IV details the system design of the Portable Tour Guide. Section V highlights system implementation and optimization, while Section VI discusses the obtained results. Finally, Section VII concludes the paper.

XXVI. RELATED WORK

Looking at the international tourism incomes in 2010 (a total of US \$927 billion) and 2011 (a total of US \$1,030 billion [3]), it can be stated that tourism is exceedingly vital to the world's economy. Many institutes have worked on developing various electronic products to serve the tourists and facilitate their tour trips. Basically, those products are divided into two main categories: General-Purpose Systems and Tourism-Dedicated Systems. These categories are described below.

A. Tourism-Dedicated Systems

Those devices are embedded systems designed for the purpose of tourism guidance only. These systems cannot be

All authors are from the Department of Electrical & Computer Engineering, Sultan Qaboos University, P.O. 33, Al-Khodh, Muscat-123, Sultanate of Oman (phone: 968-24142539; fax: 968-24144336; e-mail: amaashari@squ.edu.om, alily@squ.edu.om)

used outside the tourism site. Some examples of these systems are listed below.

1) *ARCHEOGUIDE*

Augmented-Reality based Cultural Heritage On-site Guide or ARCHEOGUIDE is a device which aims to provide a tourist with information related to cultural heritage sites. The device consists of a transparent head-mounted display, a headphone, and a portable computing unit. The device is meant to be provided only at the targeted site. This portable computing unit has got a position tracking sensor, which interacts with some elements found in the site. After the position has been identified, the portable computing unit communicates with the site's server through a wireless local network and obtains the necessary information from the database about that specific location. Finally, the information is displayed to the visitor in a user-friendly way featured with 3D-visualization in real-time. In addition to displaying the superimposed image, ARCHEOGUIDE also speaks to the visitor through the headphones.

ARCHEOGUIDE was first tested in one of the well-known European cultural heritage sites in Olympia, Greece. Later on, several sites have implemented it. The disadvantages of this system are: (1) it requires high implementation effort, (2) expensive, (3) and serves specific sites only [4].

2) *Audio Guidance Systems*

Some historical sites in the UK use audio guidance systems in their tourism sites (e.g. castles). This technology is available for all visitors. When paying a nominal fee; each visitor receives a remote controller and headphones. While wandering around the castle, the tourist will find numbered tags placed in certain rooms, corridors, and monuments. The tourist uses the remote controller to choose the corresponding tag number and listens to the educational information about it. Other tour guidance systems may be huge screens placed at the entrance of a landmark. These devices show the map of the landmark and some other related information.

3) *Prague Electronic Tour Guide*

Prague electronic guide is a device that is designed for tourist areas in Prague city, Czech Republic. It is an audio device equipped with GPS module and buttons. The embedded GPS module provides the user with audio information regarding the tourist scenes that is close to the user. These audio information is updated while the tourist is walking around the city with the help of the GPS module. In addition, the device is able to save four locations to be revisited by the tourist due to the presence of equipped navigational system [5].

B. *General-Purpose Systems*

This section presents multiple-use systems, where each system can be used for many applications including tourism guidance.

1) *Google Glass*

Google glass is a wearable device where the lenses are

interactive. Using voice commands, google glass respond to user requests and show the result through lenses. It can provide variety of services to the user such as identifying Points of Interests (POI's) that are spotted by the tourist and giving directions [6].

2) *Smartphone Applications*

Tourism applications that are used mostly by tourists are classified based on the services they provide; namely, navigation, entertainment and information. Most of the existing navigational applications, which help the tourists find their way, are either AR-based or position-tracking based. An example of AR-based navigation applications is Layer app [7]. On the other hand, entertainment applications provide the tourist with a list of directions to entertainment places POI's (e.g. cinemas, theaters, festivals if any etc.), while information applications provide information about tourism in a specified country.

XXVII. EMBEDDED VISION AND AUGMENTED REALITY

Computer vision is a discipline that is concerned with emulating the human vision. This discipline combines both image processing techniques and intelligent algorithms in order to analyze scenes. With the abundance of computational resources in embedded systems, a new technology has emerged; namely, Embedded Vision Systems (EVS) [8]. This new technology allows small, but smart, devices to analyze scenes and make decisions accordingly.

There are a number of factors that made EVS a reality, including the availability of small inexpensive HD cameras, sufficient resources on embedded systems, and the breakthroughs made in the field of computer vision. On the other hand, researchers are investigating means to overcome challenges such as power consumption and real-time experience, to list a few.

Augmented Reality, AR, is one direct application of EVS that has gained a lot of attention recently. In essence, AR refers to augmenting real-world images with computer-generated multimedia such as graphics, text, audio, and video. AR has many applications in sectors such as education, industry, and tourism. AR relies heavily on computer vision techniques such as object recognition. For example, an embedded vision system may need to recognize a landmark, before it can supply useful information to the tourist.

Object recognition finds objects in the real world from an image of the world, using object models, which are known a priori [9]. Object recognition is usually a twofold process. First, features are extracted from the image. These features are representative of the object in the scene and can be used later to detect or recognize the same object when it appears in other images. These features has to be robust; in other words they need to be invariant to illumination, orientation, scale, and even affine. The second fold of the object recognition process is where these features are either matched to model object or classified into a predefined object category.

A. *Speeded-Up Robust Features (SURF)[10]*

SURF is an algorithm that extracts features from an image. These features, referred to as keypoints, are invariant

to translation, scale, orientation, and even affine. SURF is composed of several processing stages. The first stage is concerned with building a scale-space pyramid, which is controlled by the number and size of filters that are applied to the image. SURF gives the flexibility in choosing the layers (i.e. filter sizes) and octaves (i.e. groups of layers). Generally speaking, larger number of octaves and layers yield better accuracy, however, longer execution time. The authors of SURF specify 5 octaves, each of which hosts 4 different layers. This doesn't prevent the user from trading off accuracy for speed by reducing number of octaves and/or layers.

The remaining stages in the SURF algorithm are concerned with computing the orientation of each detected keypoint and describing these keypoints with a distinguishable value. Keypoints can be described either with 64 or 128 values. Although the 128-value mode (AKA Extended Mode) yields better accuracy, however, the execution time is slower when compared with the 64-value mode.

B. Brute Force Matching

Once keypoints are extracted, then we can attempt to find a match for the object at hand. The Brute Force matching is a commonly used technique that attempts to match each keypoint in the input image to all keypoints in the database. Also, it is worth mentioning that the standard metric for finding the distance between input keypoints and model keypoints is Sum of Squared Distance (SSD).

The following section discusses how these two computer vision algorithms are employed in order to support AR on the mobile devices of the tourists.

XXVIII. SYSTEM DESIGN OF PORTABLE TOUR GUIDE

This section describes the proposed Portable Tour Guide (PTG) system. We begin by outlining the engineering requirements that were identified by the authors to achieve the system's objectives. These requirements are:

1. System should superimpose text labels describing real-world images captured by the tourist.
2. System should permit the tourist to display more details about the captured image or show a nearby Point Of Interest (POI).
3. System should be accessible and easy to obtain.
4. System should not weigh more than 300g to make it comparable to the weight of mobile devices usually carried by tourists.

After identifying the requirements and examining existing systems, the authors compared several conceptual designs. Examples of such concepts are:

1. An embedded wearable device that contains a camera unit to capture photos, a processing unit to process data and a display unit (a screen) to show results. This device can be attached to the user's body and the display is held in the user's hand.
2. An embedded device with buttons, wireless access, a built-in GPS tracker, and headphones. When the device is turned on, the user's location will be

processed. The device will receive audio information regarding the location and any nearby POI's. The user can use the buttons to turn the device ON/OFF, send location, repeat the audio information, and save audio.

3. A handheld device which points to a tagged POI to get information from it. The tag contains certain information about that POI. The device transmits electromagnetic waves that recognize the tag and takes necessary information about the POI from it.
4. An embedded device that has a camera, display, and a communication method to cloud computing service. The user takes pictures of POI and the images are sent via the communication method to the cloud for processing. The information related to that POI is sent to the embedded device for display (real-world images annotated with information).

Each of the proposed concepts has a set of advantages and disadvantages. Table 1 highlights the advantages and disadvantages of each design concept.

Table 1 Comparison between proposed design concepts

	Advantages	Disadvantages
1	<ul style="list-style-type: none"> - Allows the implementation of AR - Does not require wireless connection 	<ul style="list-style-type: none"> - Hard to update for multiple users
2	<ul style="list-style-type: none"> - Doesn't require image processing - Portable 	<ul style="list-style-type: none"> - High wireless bandwidth consumption
3	<ul style="list-style-type: none"> - Fast response - Allows the implementation of AR 	<ul style="list-style-type: none"> - Impractical to attach tags everywhere - Limited availability
4	<ul style="list-style-type: none"> - Allows the implementation of AR - Doesn't require extra components other than the ones the user owns - Easy to update - Can be implemented in real-time 	<ul style="list-style-type: none"> - Relatively high power consumption - High wireless bandwidth consumption

Based on the comparison shown in Table 1, it was found that concept number 4 is the most suitable. The most encouraging advantage of this concept is that its components (camera, communication means, etc...) are all incorporated in one embedded device. Hence, this reduces the overhead in the system because less interfacing between the components is required (unlike concept number 3).

Since, most AR existing products are smart phone applications, we have chosen smart phone to implement this system. The final product would be more appealing to the users since they are familiar with it. Moreover, choosing smartphone as the platform for implementing this system satisfies most of the system requirements. However, choosing the 4th concept brings forth two major challenges: (1) high power consumption and (2) limitations of communication bandwidth.

The proposed PTG system follows the client/server paradigm, as illustrated in Fig. 2.

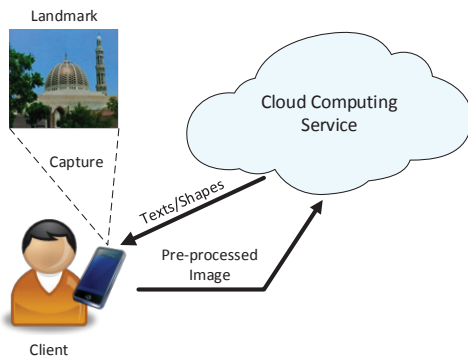


Fig. 2 PTG system adopts the client/server paradigm

A tourist (i.e. client) uses a smart portable device (e.g. smart phone) to capture a photo of the landmark that the tourist would like to inquire about. The captured image will be processed before being communicated to the cloud. In turn, the cloud will further process the received image, where extracted features are used to recognize the landmark. Once the landmark is recognized, the cloud will respond to the query of the tourist by sending descriptive text and shapes. Finally, the smart portable device will use the text and shapes to augment the captured photo as discussed earlier.

The following subsections detail the design of both client and server.

A. Software Design of the Smart Device Application

To facilitate the steps executed by the client, we propose the following process, which is depicted in Fig. 3.

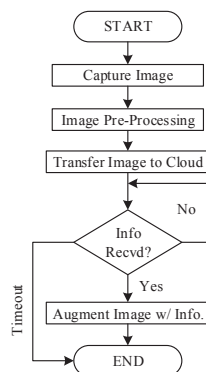


Fig. 3 Client side Process

The process starts by capturing an image frame of the landmark. Image sensors, found in smart portable devices, vary in the supported features such as frame resolution and bit representation of pixels. The software uses OpenCV library [11] to pre-process the images by resizing them to nominal frame size. In addition, the software converts the color space of the captured frame to grayscale. Pre-processing the frame serves two aspects. Firstly, making captured frames uniform in terms of resolution and color space, which helps in improving the accuracy of object recognition. Secondly, communicating captured frames in grayscale uses less bandwidth (i.e. 8-bit grayscale vs. 32-bit RGB), which results in reduced transmission delay. The software sends the processed image along with other information (e.g. image size) to the cloud.

At this point, the software awaits for the cloud to respond to the query. To avoid deadlocks, the software sets a timer such that if no response is received from the cloud, then the

client can either try again or terminate the process if the cloud is unreachable. In case the cloud do respond to the query, then the software will augment the captured image with the information founded in the response. This process is then repeated for every captured image.

B. System Design of the AR Cloud Computing

The proposed Cloud Computing System is shown in Fig. 4. The system is composed of the following components:

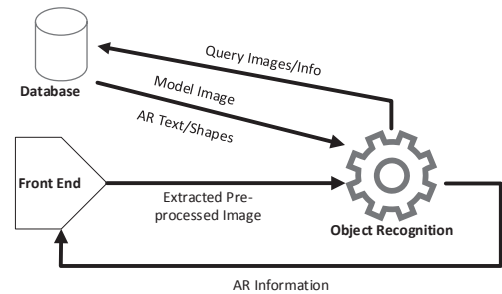


Fig. 4 The cloud computing service is composed of three components; front-end, Object recognition engine, and a database that hosts model images and landmarks information

1) The front-end

The system listens to incoming requests from clients. As soon as a proper request is received, the front-end will extract the image frame and other useful information from the payload. Then, this information is passed to the Object Recognition component.

2) Object Recognition

This component applies computer vision techniques on the received image frame. First, SURF algorithm is used to extract features from the image. Then, these extracted features are matched to pre-existing features. Once a match is found, then the landmark is identified. Then, the landmark is used to query the database for information relevant to that particular landmark. This information will be transferred to the client that made the request.

3) Database

The database is an essential component of the system. It hosts a repository of landmarks in Oman, historical information and interesting facts about each landmark, and other statistics (e.g. how many times this landmark was queried in the last month). In addition, the database implements a multi-level access system; allowing administrators to manage the content of the database.

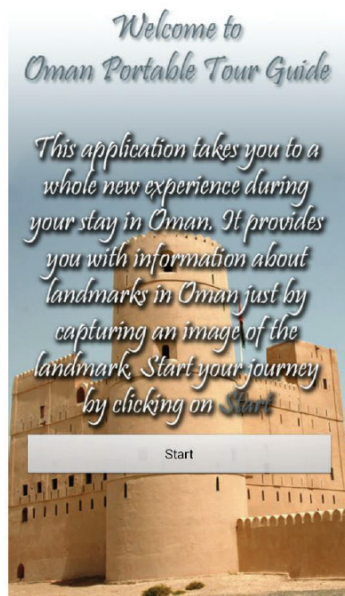
XXIX. SYSTEM IMPLEMENTATION & OPTIMIZATIONS

The proposed PTG system for tourism was developed entirely with FOSS. The following subsections outline the details of the implementation.

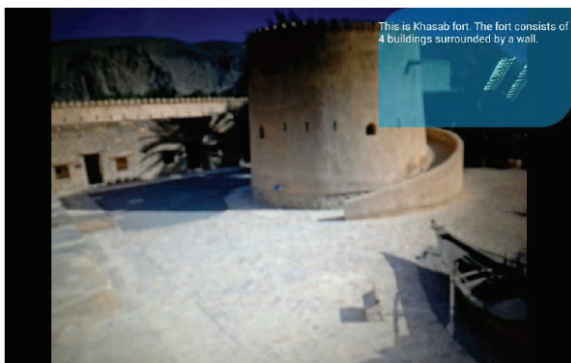
A. Client Operating System and Application

We choose Android OS [12] to operate the mobile device at the client side. This choice was made because Android is FOSS, widely used in embedded systems, and the development environment is readily available. Android SDK [13] was used to develop the application running at the client side. In addition, OpenCV for Android [14] was used for performing image pre-processing that was discussed in

Section XXVIII.A. Fig. 5 shows the user interface of the application.



(a) Welcome Screen



(b) Landmark enhanced with AR

Fig. 5 User interface of the mobile app, which is developed within Android platform

B. Server Operating System and Applications

The server is expected to be reliable and highly responsive. Linux CentOS 64-bit was used for the server platform. In addition, Eclipse IDE [15] was used for development, along with OpenCV C++ library for computer vision. In addition, XAMPP for Linux [16] was used since it combines the following package:

- **Apache server:** this webserver provides administrators a web-based interface to access the database for administrative purposes.
- **MySQL server:** this server facilitates database management.
- **PHP and Perl:** these two packages are mainly used for web development.

We anticipate that the landmarks repository will grow in the future. Hence, the database needs to be expandable. To facilitate that, a web-based interface is implemented for administrators; allowing them to easily upload new landmarks, edit existing ones, or delete a specific landmark.

C. System-level Optimizations

It is expected that as the number of tourists that use the proposed system increase that the server will be hammered with large number of client requests. In addition, as the size of the landmark repository increases, the matching stage in the Image Recognition component will dominate the execution time. This is because the server will attempt to match the features extracted from the query image to all features extracted from all model images. Therefore, it is crucial to design a scalable system to cope with such increase in demand.

In the last decade, we have witnessed an abundance of parallel processing capabilities, especially in the Many Integrated Core architecture (MIC) [17, 18] and Warehouse-Scale Computers (WSC) [19, 20]. The use of these parallel resources is the solution to develop a scalable system. Therefore, we have used the POSIX library to parallelize the workloads across the available processing cores. First, feature extraction of the query image was performed on a single core. Then, the extracted features are distributed across available cores. Each core has a subset of the model images' keypoints. Therefore, each core performs matching on a subset of the model landmarks, and hence reducing the total execution time for matching.

XXX. DISCUSSION OF RESULTS

The proposed PTG system was prototyped and tested for execution time and accuracy.

A. Experiment Setup

For the client side, the developed application was installed on a Samsung Galaxy S4 smart phone [21]. The phone has 1.9GHz quad-core processor, 2GB RAM, and is operated by Android 4.2.2 "Jelly Bean" OS.

The server side was hosted on an Intel quad-core i7 machine clocked at 3.4GHz. The machine supports 64-bit architecture and contains 8GB RAM.

B. Design Space Exploration

This subsection provides an exploration of the design space. Since the machine vision algorithms dominate the execution time, we narrow down the dimensions of the design space to a number of octaves and layers implemented in the SURF algorithm.

Fig. 6 shows the relationship between number of SURF octaves and layers from one side, and execution time and accuracy, on the other side. The figure indicates that accuracy is improved by increasing number of layers and octaves. On the other hand, the execution time will also increase as a result of more computational load when increasing number octaves and layers. To balance between accuracy and execution time, we find that the optimum number of octaves and layers is 4 and 2, respectively.

C. Accuracy and Execution time

Next, we experiment with the system using the design configurations outlined in the previous section. We use an image dataset that is composed of 220 images for a total of 46 different landmarks. All images have a size of 800×600, since this size retains the main features of a landmark.

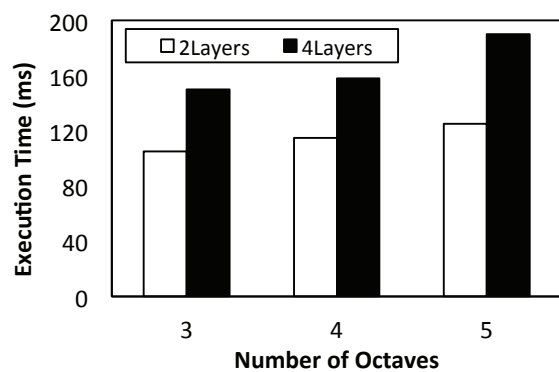


Fig. 6 Execution time of SURF algorithm using different configurations

Experiments reveal that on average, the total execution time is 936ms, where 62% of the execution time is dominated by the communication delay. On the other hand, the pre-processing that takes place in the mobile application (i.e. from the time an image is captured until the grayscale version of the image is packaged and ready for transfer) contributed by 78ms, while processing the image at the cloud took 278ms seconds.

We expect that the execution time at the client side will diminish since more powerful processors are being utilized in mobile devices and embedded systems. Similarly, the communication delay can be reduced by transferring compressed images that utilizes lesser bandwidth.

D. Parallel Processing

To reduce the execution time at the server side, we parallelize the matching process across the cores. Fig. 7 shows how increasing the number of utilized cores reduce the execution time. For example, when utilizing 8 cores, the overall execution time of object recognition is speeded up by 12.5%.

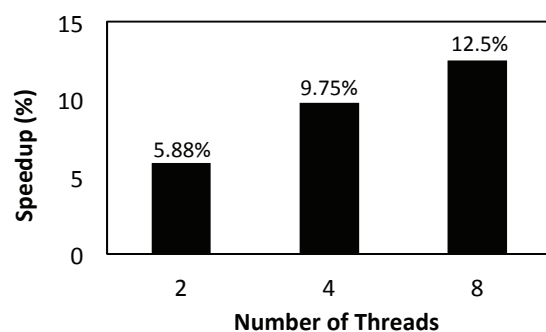


Fig. 7 Speedup in execution time when parallelizing the brute-force matching stage across multiple processing cores. Results are normalized to 1-core execution time.

XXXI. CONCLUSION & FUTURE WORK

This paper presented a proposed system that targets the Omani tourism sector. The proposed system was developed and implemented purely by FOSS software. The system was effectively deployed inside the campus of Sultan Qaboos University in May 2014.

For future work, we are planning on making further optimizations to improve the performance and accuracy of

the landmark recognition. For instance, the system can make use of the GPS information to narrow down the number of points of interest (POI) to search for. Additionally, Object classification using Support Vector Machine (SVM) will be investigated.

REFERENCES

- [1] K. Yousuf. "Needed: More tourist guides," October, 2013; <http://main.omanobserver.om/?p=24541>.
- [2] T. f. Oman. "Oman set for 41,000 jobs in tourism in 2014," <http://www.timesofoman.com/News/33435/Article-Oman-set-to-lead-the-Middle-East-tourism-sector-in-2014>.
- [3] W. T. Organization. "UNWTO Tourism Highlights," https://s3-eu-west-1.amazonaws.com/storageapi/sites/all/files/docpdf/unwtohighlights12_enlr_1.pdf.
- [4] N. I. Vassilios Vlahakis, John Karigiannis, Manolis Tsotros, Michael Gounaris, Didier Stricker, Tim Gleue, Patrick Daehne, Luis Almeida, "Archeoguide: An Augmented Reality Guide for Archaeological Sites," *Computer Graphics in Art History and Archaeology*, 2002.
- [5] E. T. Guide. "Prague Electronic Tour Guide :: about GPS audio-guide," <http://electronicguide.com/>.
- [6] Google. "Google Glass," <http://www.google.com/glass/start/>.
- [7] A. R. Blog. "Augmented Reality," <http://nipunsangral.blogspot.com/>.
- [8] E. V. Alliance. "What is Embedded Vision," <http://www.embedded-vision.com/what-is-embedded-vision>.
- [9] R. K. Ramesh Jain, Brian G. Schunck, *Machine Vision*: McGraw-Hill, Inc., 1995.
- [10] A. E. H. Bay, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.
- [11] G. Bradski, "The OpenCV Library," *Dr. Dobbs Journal of Software Tools*, 2000.
- [12] Google. "Android," <https://android.com/>.
- [13] Android. "Android SDK | Android Developers," <https://developer.android.com/sdk/index.html?hl=i>.
- [14] OpenCV.Org. "Android | OpenCV," <http://opencv.org/platforms/android.html>.
- [15] T. E. Foundation. "Eclipse," <https://eclipse.org/>.
- [16] A. Friends. "XAMPP Installers and Downloads for Apache Friends," <https://www.apachefriends.org>.
- [17] R. Wilson. "From Multicore to Many-Core: Architectures and Lessons," <http://www.altera.com/technology/system-design/articles/2012/multicore-many-core.html>.
- [18] J. Jeffers. "Intel Many Integrated Core Architecture: An Overview and Programming Models," https://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/ORNL_Elec_Struct_WS_02062012.pdf.
- [19] D. P. J. Hennessy, *Computer Architecture: A Quantitative Approach*, 4th ed.: Morgan Kaufmann, 2012.
- [20] U. H. Luiz André Barroso, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*, 2009.
- [21] Samsung. "Samsung Galaxy S4," http://www.samsung.com/ae/consumer/mobile-phones/smart-phone/android/GT-I9500ZWAXSG?cid=ae_en_mob_s4_aon_ggl_ppc_dm_text_exact_2_0140918.