

An Augmented Reality Multimedia (ARM): A Dynamic Multimedia Encyclopedia on Mobile Device

N.Venkatesvara Rao, D.V.V.Prasad, Rajesh. M

Abstract: *In the recent days, there are rapid advancements in the web 2.0 contents (texts, images and videos) and in the computational complexity. The next generation in computing is predictably in the mobile arena. A lot of applications are being developed to cater the needs of mobile device users. This research is based on the question: How can we benefit from the rapidly proliferating web 2.0 contents in the mobile device with staggering technological developments? This paper presents a surprisingly impressive novel dynamic multimedia encyclopedia, multimedia, that augments reality on the mobile device. This system supersedes the traditional texts or maps retrieved. A group of estimation and sensors to track objects and keep information on them used. An augmented reality based multimedia encyclopedia on mobile device which is dynamic is presented. It is dynamic and mines resources on the web. We demonstrate how this device can be utilized to bring relevant information to users in a much more enriched manner than traditional texts. ARM is dynamic and is distinctive in 3 ways. It fetches multimedia contents for users through automated intuition and dynamically updates content from the web. It uses recent multimedia content to produce answers which are versatile. New mediums like videos & languages can be incorporated into the framework. This system takes advantage of the mobile camera, location and orientation sensors to augment reality. Live camera view and retrieve dynamic data. A Combination of algorithms is used to analyze and demonstrate our ARM.*

Keywords: *multimedia, encyclopedia, mobile device, augmented reality*

I. INTRODUCTION

Augmented reality (AR) is the real time direct or indirect lead view of the surroundings whose characteristics are transformed or supplemented by system generated sensory inputs from various sensors like GPS, orientation sensors, and camera. AR technology makes it possible to interactively and digitally manipulate the environment with these inputs to generate artificial information in a multimedia format. A systematic search

on any search engine would reveal the recent interest and advancement in AR. In the recent years a lot of methods and protocol implementation are suggested to users through mobile internet devices. Sophisticated algorithms have in the recent years to accurately augment the real time feed. The application of AR spans a variety of field like archaeology, architecture, art, commerce, construction, Education, Gaming, Medical, Military, Tourism, Translation and Navigation. However, ours is the first attempt to amalgamate multimedia mining of web 2.0 and augmented reality on a mobile device. In this paper, we propose a multimedia encyclopedia called multipedia which automatically retrieves, updates by utilizing the rich resources of web 2.0 that utilizes the augmented data of the real world environment on a mobile internet device. This is the era of smart phones and most smart phones or Mobile Internet devices (MID) with WLAN, connectivity, Global Positioning System (GPS), orientation and motion sensors. The contemporary mobile applications (apps) for AR depend only on the location orientation sensors. Our system extrapolates the contemporary by matching algorithms for the real generate a data. The augmented data is then subjected to a novel implementation, multimedia mining to create a multimedia encyclopedia, multipedia on the mobile device, which gives users a new experience to interact with the real time environment and retrieve data about it. We use sophisticated algorithm for the web 2.0 mining and integrating it on the mobile device. An encyclopedia is a compendium of reference work which caters a summary of information about a subject. It differs from dictionary in catering the subject, instead of a word [1]. The word encyclopedia was first used as a noun by a creation encyclopedist parao skalic in his encyclopedia seu orbis disciplinarum tam sacrarum quam prophanarum epistemon (Encyclopedia, or knowledge of the world of Disciplines, Basel, 1559). The term as a title of a book is first used by bachimus forties Ringelbergies in the year 1541 [2]. The encyclopedic entries are longer and more detailed than dictionaries [3].

Thus encyclopedia is a means to disseminate knowledge around the globe and to transmit information. In the advent of latest technologies since 1985, Encyclopedias found their way

Revised Manuscript Received on September 22, 2019.

N.Venkatesvara Rao, Saveetha school of Engineering , Saveetha institute of medical and Technival sciencea raosusipooji@Gmail.com

Dr.D.V.V.Prasad, Professor in CSE,SSN College of Engineering, dvvprasad@ssn.ac.in

Rajesh. M, KRS College of Engineering, Vandavasi, India, Raga Academic Solutions, India

in CD-ROMS supplemented by visual information. The multimedia revolution of 20th century brought videos, audios and high definition images to supplement these digital encyclopedias. In the recent years web based encyclopedias such as Wikipedia, Encyclopedia Britannica, subject specific encyclopedias such as FARS, Botanicus bloomed due to the web 2.0 revolutions. Image sharing sites (web 2.0) like flickr, deviantArt, zoomr, fotki, pBase.com, photoBucket, photo.net, smugmug, 500 px and Instagram allows anyone to share images, distribute, evaluate and interact through social networks like LinkedIn, facebook and google plus. The images allows user to geotag them with the geographical information or location, which allows the shared images to be mapped for exact location through technologies like Global Positioning System (GPS) later when shared. YouTube offers videos to be shared and interacted. By the availability of image, text and video encyclopedia which can be accessed through web, it is possible to use these valuable resources to augment the reality. [4] [5].

Recently, a variety of methods have been suggested to present augmented through mobile devices [6] [7]. Most smart phone which are available now are equipped with High Definition cameras, WAN, WLAN and Global Positioning System [8]. Projects such as Encarta, Wikipedia and others include considerable amount of images and videos. However, they are not fully fledged with cameras, WAN, WLAN & GPS integration to produce augmented reality encyclopedia. In this paper, we propose a multimedia encyclopedia called Augmented Reality Multipedia (ARM) which integrates web 2.0 encyclopedia with the features of the modern mobile device to produce a novel encyclopedia. Our system aims to use image matching protocol to recognize objects to produce augmented reality. We implement this in a latest mobile internet device (MID) with the aforementioned features. The key characteristics of ARM that distinguish it from others are 1. The output is an augmented reality video 2. It is dynamic, as it is updated with information via the integrated features. It is completely versatile and extensible.

II. SYSTEM OVERVIEW

The following are the system construction overview dealt in this paper. a) Location and orientation sensors are utilized to track the objects.[9]. b) Creation of sampled image database from the large database to test MAR performance in a realistic setting. c) Image Clustering for producing exemplars. d) Discussion of the association of exemplars with the database. e) Design of an identical algorithm based on histograms of minimum distances between feature descriptors and furnish results from our database [mar]. f) Using GPS, Google image search and Wikipedia to present and extend our database, thus enhancing the image matching of our system. g) Assembly of Exemplars to produce image encyclopedia (multipedia) where the augmented image is transcript in a vivid way. The organization of this paper proceeds as follows: In part 2, we give an overview of our augmented reality system. In part,

we represent a sampled database. In part 4, we propose image clustering for producing exemplars. In part 5, we present an identical algorithm to produce results from our database. In part 6, we enhance our database by images from Google image search and better matching results. In part 7, we illustrate the implementation of our proposed multipedia through augmented reality. Finally we conclude in part 8.

Part 2: The AR System Illustration

The Augmented reality system is based on two components. A server and a client Mobile Internet Device (MID). The MID communicates with server through WAN or WLAN network. We demonstrate and implement a complete end to end mobile augmented reality multipedia (ARM) on a device that has an intel atom processor and web based - MAR service hosted on a server, the server on which the MAR service is hosted is updated on a regular basis with the web 2.0 contents. The MAR client application is running on a MID. Fig 1 demonstrates the actual snapshot of the client interface. The example picture, we have taken to demonstrate is an iconic image in Chennai, Tamil Nadu, India called Anna arch. Our system uses the location data along with the camera image to hit the database to retrieve the top hits of the image server. The user can interactively select the image of interest and retrieves the web 2.0 contents relevant to the interest. The retrieval of web 2.0 contents is directed by our algorithm which is implemented in the following steps, which are explained in the later part of this paper. 1. Clustering the images for creating exemplars 2. Relating such exemplars to the web 2.0 against noise tag filtering 3. Presenting the encyclopedia on the mobile client interface. 4. Evaluation of the multipedia on the client.

2.1. The Client:

a. Acquisition of Query:

The camera, orientation sensors and location sensors like GPS are used continuously to acquire live video from the camera in the client MID. As soon as the picture is selected by the user and taken, the sensor and location data are written to the image EXIF fields.

b. Feature Extraction:

Surf [7] is used to extract the features of the picture and it is 64 dimensional. The feature vectors and sensor data are sent to the server or database by the client to lookup for matching images and relevant information. Network bandwidth and latency are reduced by searching with the features of the images instead of high resolution pictures.

c. Augment Generation and overlay:

The matched information is downloaded through WAN or WLAN and tag of the query object is overlay on the live feed. The overlay is the augmented information on the live feed which uses compass,

accelerometer, and gyroscope and pinned to the feed. The augmented information can be viewed by simply pointing the camera at different places.

d. Trailing (Tracking):

The client continuously tracks orientation sensor and the directions of MID are recorded when query is given [6]. Since the methodology presented in [6] is not accurate. We use a motion estimation algorithm [9] which is based on iterative gradient based strategy. It is statistically complex the two different motion mode is used in this strategy are pure camera rotation, which has 3 parameters and pure translation which has 2 parameters [9].

2.2. The server

a. The Database:

As per [6], the pages relevant with GPS information were crawled and the images were downloaded from the pages to the sever. The Visual features of such images were extracted and build in our image database. The exponential growth potential of the database suggests a huge arena for exploration in the image resources domain.

b. Matching:

The Features of the image were sent to the server from the client MID. For the purpose of demonstration, we restricted the search to the database with in a particular GEO location using the user GPS data. The GPS influenced data set is then subjected to matching using image based matching algorithms like brute force image matcher [8], which relies on finding the match by deducing the distance between nearest and second nearest neighbor descriptor for such data, which is without the GPS data, indexing methodology were handled [1]. Details of the algorithms are furnished later. The combined effort to produce our ARM is explained in the following algorithmic stages.

1. Clustering the images for creating Exemplars:

Exemplars are imitations or similarities which are pivotal in effectively organizing images for their attributes. Clustering or grouping image attributes for enhanced user experience were suggested by [1] [3]. But these works are traditional in the way it organizes clusters and challenges do exist in the way information is presented to the users [3] uses k-means clustering. Frey et al formulated Affinity Propagation (AP) to discover similarities (exemplars) through data points and thus effective than [1] [3]. The AP algorithm is used to get exemplars for representations through data points in images. From the studies of [12] and [5] we know that AP algorithm is an effective measure to find image exemplars. 'n' data points are represented as $X = \{x_1, x_2, \dots, x_n\}$. The exemplarity (similarity) between two data points is given by $E(x_i, x_j)$. The data points are assembled into m clusters, where each group is denoted by an exemplar element from X, where $(m < n)$,

There are Q different types of information that are implemented through AP algorithm

1. Responsibility R
2. Availability A for all (i,j). The in formations are given in the following equations. [main]

$$R(i,k) = \max_{(k' = k)} \{A(i,k') + E(x_i, x_{k'})\} \quad ; \quad (1)$$

$$A(i,k) \leftarrow \min\{0, R(k,k) + \sum_{i' \neq k} \max\{0, R(i',k)\}\} \quad (2)$$

$i' \in \{i, k\}$ Where,

$A(k,k)$ Denotes self availability and given by

$$A(k,k) \leftarrow \sum_{i' \neq k} \max\{0, R(i',k)\} \quad (3)$$

The exemplar for each data point x_i is $e(x_i) = x_k$ where k maximizes the following

$$\arg \max_k A(i,k) + R(i,k) \quad (4)$$

2. Consorting of exemplars to the web pages

The AP algorithm promotes the determination of exemplars for each cluster. Having the exemplars at hand, the next difficulty is to find where and how to get the accompanying text description and weed out the unnecessary hits (pruning) for those exemplars. Most images which are uploaded online contain a huge number of user tags in social media networks. These tags contain noisy information, which are not useful to understand the characteristic meaning of the images. The noises should be eliminated first and the relationship of the rest of the tags with the individual exemplar is correlated to the corresponding webpage. By summarization technique the web pages were pruned.

3. Filtering the noisy tags

Our objective is to detail the exemplars with their accompanying text descriptions in the web pages. The idea is to leverage on the tags of the exemplars. Hence the quality of such tags needs to be high. Tags like Typo, number, model ID and stopwords are irrelevant and must be eliminated from the tag list. As suggested by [8], we used a lexical database and removed noisy tags through wordNet. The tags in the respective word group are listed after removing unnecessary tags that are not listed in wordNet. We use Normalized Google Distance (NGD) [10] to attain a semantic relation between the concept and its accompanying tags. NGD is a metric for determining semantic interrelatedness obtained from the number of search results returned by Google search engine. For a concept q and tag t_{ij} , the NGD is given as follows,

$$Ngd(q, t_{ij}) = \max\{\log f(q), \log f(t_{ij})\} - \log f(q, t_{ij}) / \log M - \min\{\log f(q), \log f(t_{ij})\}$$

Where,



M – Total number of retrieved web pages.

$f(q)$ – number of search results for the concept

$f(t_{ij})$ – number of search results for tag

$f(q, t_{ij})$ – Number of web pages having both concept and tag.

The tags were ranked with respect to the concept based on the NGD score. AP algorithm clusters only the images with feature similarity, hence some tags may be found to be existing in more than one cluster.

4. Linking

For the concept we are handling, we obtained many exemplar images via clustering the retrieved images from an image database. The corresponding textual documents are collected from the internet. The combination of exemplar images of the query and the textual descriptions from the internet could provide a good understanding to the users [15], but the task is extremely difficult to automatically link each image with the document. We turned this problem into a computational issue. We utilized the Latent Semantic Analysis (LSA) algorithm [14]. LSA is a method for automated indexing and retrieving information which maps the textual document and represents them in Latent Semantic Space. Hence, it is a protocol for us to determine the similarity between individual image and text documents and the score of the LSA the similarity score is used to link the image and the descriptions. The detailed version of LSA can be found in [14] for the interested reader. Hence we give the relevant equation alone to represent our for a concept q consisting of the tags of a image, its accompanying representation in k -dimensional latent semantic space is $q^T u_k$ and the textual descriptors are represented as columns of $\Sigma_k v_k^T$. The similarity between our concept and the textual document is given by the score, $\text{Sim} = (q^T u_k) (\Sigma_k v_k^T)$

5. Presenting the encyclopedia through an interface:

Upon associating the concept and documents through exemplars, the next task is to present the information in an elegant and coherent way to the user, so that it is visual appealing though there are numerous. Summarizing protocols for individual and multiple web documents. These methods are improvised by relying on improvements like hypertext structure including anchor text, and web framework. [6] Proposed a model to summarize by utilizing concepts and inforbox structure through the extended document concept lattice model and provides customizable detailed information to suit users needs. To present the summary in a visually enhanced way, we deployed. API from image loop [8] for presenting our multimedia encyclopedia through slide show which presents image according to the size of the clusters. The user interface is accompanied with informations that is presented with a background music.

6. Constructing a database

For a better database, we selected ten landmarks in and around chennai and for each individual places, we

followed our protocol in our mobile augmented reality system. We refined the images database on the server to a lesser set that contained images within 10kms. We selected one landmark image as query and stored the others as database images. We determined the number of relevant images out of the total number of images placed in the database. We curate the database manually to attain quantitative results denoting the database performance since our database is a sampled database.

7. Matching Improvisation algorithm

To identify the most matching database images, our algorithm checks individual database images one at a time and equates it to the query image [8]. Upon finding the matching pairs, it is ranked in descending order. We ran the collected data on MAR and found multiple matching key points in the query image represents to one keypoint in the database. Hence we deployed duplicate removal through our algorithm to enhance the matching accuracy. This is done because, false or duplicate matches are enhanced by their increase in number which will finally affect the overall retrieval efficiency. For individual key points in the query image, the algorithm calculates a less distance and a second less distance between its descriptor and the descriptors of the database key points []. The ration is used to determine the match. We relied upon other statistics also to determine distance, to check the similarity and variance between the query and database images. We have done this through histograms to compute the distance. The algorithm is explained. Q is the query image and D is the notation representing database image. The matching algorithm is used to get top 10 database images; represented by D_1, D_2, \dots, D_{10} . To construct a histogram (Q, D_i) pair is considered and the minimum between them. The empirical mean M_i is given by

$$M_i = \frac{1}{n} \sum_{j=1}^n H_{i,j} \text{ and the skewness } S_i \text{ of the histogram}$$

representing minimum distance is given by

$$S_i = \left\{ \frac{1}{n} \sum_{j=1}^n (H_{i,j} - M_i)^3 \right\} / \left\{ \frac{1}{n} \sum_{j=1}^n (H_{i,j} - M_i)^2 \right\}^{(3/2)}$$

If the S_i value is less, the value of H_i is closer to the symmetrical shape. To eliminate the database images with very low S_i value, the following assumptions were considered in the algorithm. If S_i is closer to zero, then the histogram H_i is almost symmetric and If M_i is large; Q and D_i are considerably distant and hence, most probably mismatched. We used K- means clustering for bringing for M_1, M_2, \dots, M_{10} into two clusters. The images were removed if the cluster is found to exist with higher mean.

8. Improving matching by database expansion

The important attribute that affects the improvement in matching images is not only the



algorithm but also the size of the matching database utilized. We expanded the database by including GEO tagged, GPS location information based data to be associated with the images and WebPages to improve the efficiency of the MAR system.

III. CONCLUSION

In this paper, a rapidly proliferating web contents in the mobile device is discussed. An impressive novel dynamic multimedia encyclopedia that augments reality on the mobile device is presented. The developed system supersedes the traditional retrieved texts or maps. A combination of motion estimation algorithms & orientation sensors to track the objects and place augmented information on them is used. The system is dynamic in the sense that it retrieves resources from the web. This device can be utilized to bring relevant information to the users in a much more enriched manner than traditional texts. ARM is dynamic and is distinctive in 3 ways. It fetches multimedia contents for users through automated intuition and dynamically updates content from the web. New mediums like videos & languages can be incorporated into the framework. The above encyclopedia can be further enriched by optimizing the query used to retrieve the relevant information.

REFERENCES

1. Technique for associative learning information transclusion by Wiki-style mashups Tosic, M. Manic, M. e-Learning in Industrial Electronics (ICELIE), 2011 5th IEEE International Conference on DOI: 10.1109/ICELIE.2011.6130022 Publication Year: 2011 , Page(s): 38 – 43
2. An image-map system with spatio temporal color process functions Mori, M. ; Sasaki, S. ; Kiyoki, Y. DOI: 10.1109/ICTKE.2012.6408537 Publication Year: 2012 , Page(s): 104 – 111
3. An study of fusion processings for multimodal image retrieval Csurka, G. ; Clinchant, S. Content-Based Multimedia Indexing (CBMI), 2012 10th International Workshop on DOI: 10.1109/CBMI.2012.6269843 Publication Year: 2012 , Page(s): 1 – 6
4. A Novel Cross correlation and cross-media retrieval Xiaohua Zhai ; Yuxin Peng ; Jianguo Xiao Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on DOI: 10.1109/ICASSP.2012.6288383 Publication Year: 2012 , Page(s): 2337 – 2340
5. Sivanesh kumar, A and Khanna V, “Adaptive energy-efficient Clustering for multichannel MAC protocol” International journal of Control Theory and applications, Volume 09, Issue 36, January, 2016
6. framework e secure and streamlined data for the Semantic and mobile Web Nagar, A. Future Generation Communication Technology (FGCT), 2012 International Conference on DOI: 10.1109/FGCT.2012.6476578 Publication Year: 2012 , Page(s): 208 – 212
7. Scalable Retrieval with Sparse Projection Learning and Pseudo Mining Guan-Long Wu ; Yin-Hsi Kuo ; Tzu-Hsuan Chiu ; Hsu, W.H. ; Lexing Xie MultiMedia, IEEE Volume: 20 , Issue: 3 DOI: 10.1109/MMUL.2013.13 Publication Year: 2013 , Page(s): 47 - 57 Cited by: Papers (2)
8. .Analysis of user interests in context of Web 2.0 technologies Sommer, K. ; Braun, S. ; Gramss, D. ; Vogel-Heuser, B. DOI: 10.1109/EduCon.2013.6530091 Publication Year: 2013 , Page(s): 94 – 99
9. Detection and Automatic identification Genre of audiovisual documents Fourati, M. ; Jedidi, A. ; Gargouri, F. DOI: 10.1109/AICCSA.2014.7073235 Publication Year: 2014 , Page(s): 466 – 469
10. FoP: Never-Ending Learner for Multimedia Knowledge Extraction Subercaze, J. ; Gravier, C. Volume: 2 DOI: 10.1109/WI-IAT.2014.134 Publication Year: 2014 , Page(s): 459 - 466