VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“Jnana Sangama”, Belgaum-590 018.

PROJECT REPORT

ON

“OBJECT DETECTION AND VISUAL INNOVATION USING AR”

Submitted in Partial fulfilment for the VII Semester, BE, Information Science & Engineering.

Submitted by:

MEGHAVARSHINI M	**(1GA15IS021)**
RAHUL BHAT	**(1GA15IS032)**
VAISHNAVI BHAT MARAKINI	**(1GA15IS049)**

Under the Guidance of:

Dr. GANGA HOLI

HOD Dept. of ISE

2018-2019

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

GLOBAL ACADEMY OF TECHNOLOGY

Ideal Homes, RR Nagar, Bengaluru-560098

(Affiliated to Visvesvaraya Technological University, Belgaum and Approved by AICTE, New Delhi, NAAC Accredited with ‘A’ Grade)

GLOBAL ACADEMY OF TECHNOLOGY

Ideal Homes, RR Nagar, Bengaluru-560 098.

(Affiliated to Visvesvaraya Technological University, Belgaum and Approved by AICTE, New Delhi, NAAC Accredited with, ‘A’ Grade)

Department of Information Science and Engineering

CERTIFICATE

Certified that the project work entitled “**OBJECT DETECTION AND VISUAL INNOVATION USING AR”** carried out by **MEGHAVARSHINI M** USN **1GA15IS021, RAHUL BHAT** USN **1GA15IS032, VAISHNAVI BHAT MARAKINI** USN **1GA15IS049,** bonafide students of **GLOBAL ACADEMY OF TECHNOLOGY**, is in partial fulfillment for the award of **Bachelor of Engineering in Information Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the year 2018-19. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of Phase1 project report prescribed for the said Degree.

Signature of the Guide	Signature of the HOD	Signature of the Principal
Dr. Ganga Holi	**Dr. Ganga Holi.**	**Dr. Rana Pratap Reddy N**
Prof. & Head, Dept. of ISE	Prof. & Head, Dept. of ISE	Principal
GAT, Bengaluru.	GAT, Bengaluru.	GAT, Bengaluru.

External Viva

Name of the Examiners **Signature with date (Department Seal)**

1. ---------------------------- ----------------------------

2. ---------------------------- ----------------------------

ABSTRACT

Augmented Reality (AR) is a technology which combines virtual objects and real-world environments. Technologies like Computer Vision and Object Recognition can be used with AR to create an interactive and enhanced user experience of the real world .

The proposed system uses AR to leverage the increased computing power of smart-phones and other visual interface to build a system where the user who scans the objects through the phone or a visual interface enters the augmented reality phase and can simulate or move the entities to represent a virtual world and the moves are as per the user's choice and gets the details or the description about the object which is detected, and link related to the object is displayed through which user can know more about the object. It superimposes a computer-generated image on a user's view of the real world, thus providing a composite view. It brings out the digital world into the persons perception of the world through other dimensions as natural parts. where proposed system has used Deep learning architecture to solve the detection problem for augmented reality (AR).

This as a phase one project aims to build a module for entertainment and visual innovation for augmented reality.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our effort with success.

We are grateful to our institution, **Global Academy of Technology**, with its ideals and inspirations for having provided us with the facilities, which has made this project a success.

We earnestly thank **Dr. Rana Pratap Reddy,Principal ,Global Academy of Technology** for facilitating academic excellence in the college and providing us with the congenial environment to work in, that helped us in completing this project.

We wish to extend our profound thanks to **Dr.Ganga Holi, Prof. & Head, Department of Information Science & Engineering, GAT,** for giving us the consent to carry out this project and we would like to express our sincere thanks for being our internal guide, for her able guidance and valuable advice at every stage, which helped us in the successful completion of the project.

We owe our sincere thanks to our project coordinator, **Ms. Sharmila Chidaravalli, Assistant professor, Department of Information Science & Engineering, GAT,**, for her immense help during the project and also for her valuable suggestions on the project report preparations.

We would like to thank all the teaching and non-teaching staff for their valuable advice and support.

We would like to express our sincere thanks to our parents and friends for their support.

Meghavarshini M

Rahul Bhat

Vaishnavi Bhat Marakini

INDEX

LIST OF FIGURES

FIGURES	Page No.

LIST OF TABLES

TABLES	Page No

CHAPTER 1

INTRODUCTION

1.1 Preface

Augmented Reality is the integration of virtual 3D objects with the real environment in real time, being a variant of Virtual Reality. The foundations of this technology were established in the 1960s. It is an interactive experience of a real-world environment where the objects that reside in the real-world are "augmented" by computer-generated perceptual information, sometimes across multiple sensory modalities, including visual, auditory, haptic, somatosensory, and olfactory.

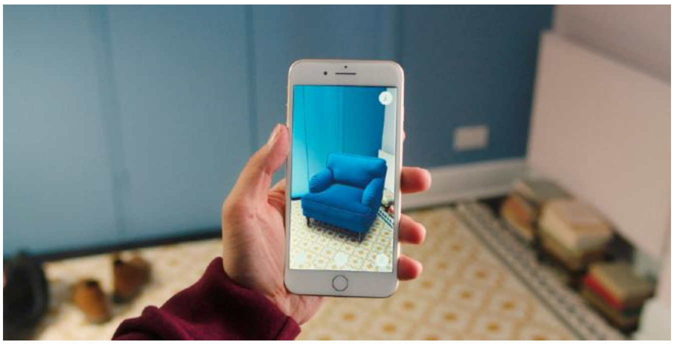


Fig 1.1 Augmented reality

Object detection refers to an ability to identify the form and shape of different objects and their position in space caught by the device's camera. Augmented reality is the enhancement of the view of the real world with CG overlays such as graphics, text, videos or sounds, and across all AR applications, object recognition is particularly severe. Most of these apps are marker-rich, which means they use special images, pictures, or objects to trigger pre-defined 3D visualization, animation, video, or soundtrack. In other words, they use object detection and tracking to determine what relevant information should be added to the real world.

Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. "Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding."[9] As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.[10] As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems

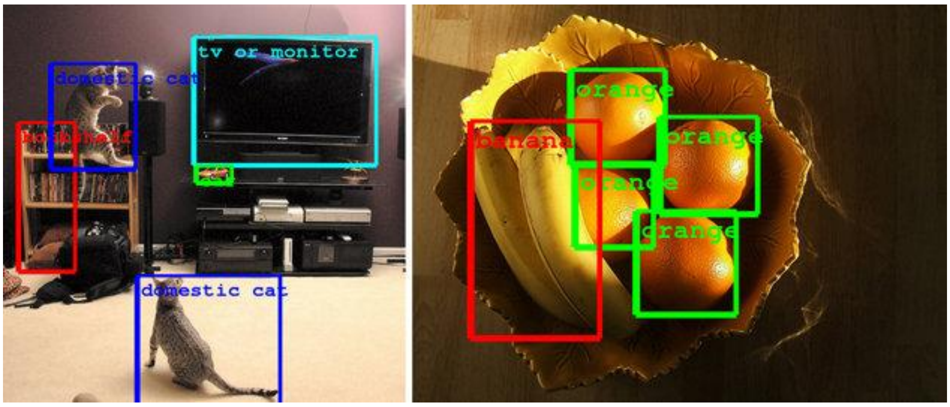


Fig 1.2 object recognition

Deep learning architecture has been adapted to solve the detection problem for augmented reality (AR). We plan to use AR to leverage the increased computing power of smart-phones and other visual interface to build a system where the user who scans the objects through the phone or a visual interface enters the augmented reality phase and can simulate or move the entities to represent a virtual world and the moves are as per the user's choice.

This project aims to detect objects in 3D space and display or link to relevant information. The project in the further stage shall implement a feature where it can be used for object tracking and navigation or gaming.

1.2 Existing System

.Vuforia

Vuforia is a standalone library that allows applications to recognize images, boxes, cylinders, text, and arbitrary objects in the environment. This tool is extremely fast, robust, and is easy to use. Moreover, it is extremely well-integrated into Unity, a leading development engine, since Unity and Vuforia have developed a solid partnership and have a shared R&D lab. Together, they introduce an irrepressible power to create amazing augmented reality experiences. Vuforia matches images caught from a camera with a pre-defined reference image. Because both images are byte arrays, searching for similar elements between the reference image and the image presented is a cumbersome task. One of most prominent Vuforia advantages is that this tool analyzes both images while searching for specific feature points.

.

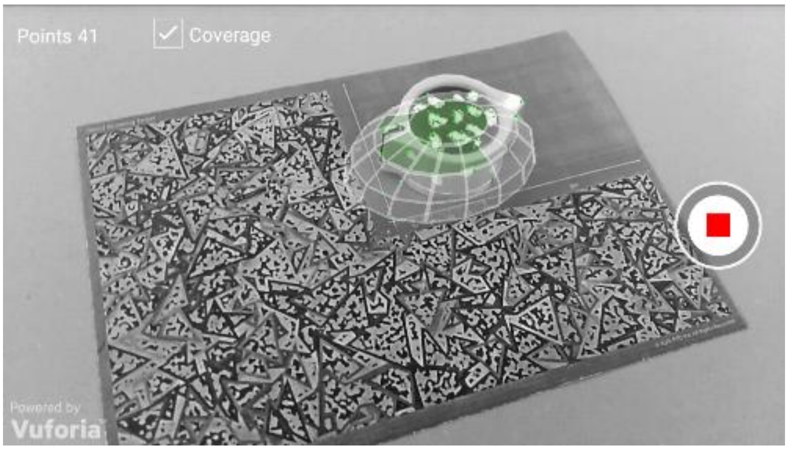


Fig 1.3 : Vuforia object scanning

Vuforia Advantages

1. Apart from being easy to master and even more easy to use, Vuforia is characterized by good integration into Unity.

2. Equally good in terms of recognizing flat, convex, and volumetric objects.
3. Markers to trigger augmented reality may be complex as well as simple.
4. Vuforia helps facilitate and simplify the development process.

Disadvantages of Vuforia

However, Vuforia is not all roses - there are many limitations. Most of them are related to flat images that act as markers and object recognition itself.

For flat markers:

1. In order to be recognized, flat markers need to have contrasting colouring, preferably with small details that will act as unique elements to be bound to.
2. For seamless recognition, markers should be matte in order to exclude flares.
3. Markers with striped elements or fully stripped, in turn, will not be recognized in any case.
4. If there is an object that covers a part of the marker, it might not be recognized.

For objects (in a flesh):

1. The chosen object has to be convex and preferably of uniform shape (without prominent protrusions), otherwise, forget about scanning it with Vuforia.
2. It is better for an object to have enough contrast points. In other words, it has to be colourful with many small details.
3. There is a maximum of two objects to be recognized simultaneously.

CHAPTER 2

LITERATURE SURVEY

Below are some of the papers which are referenced for our project

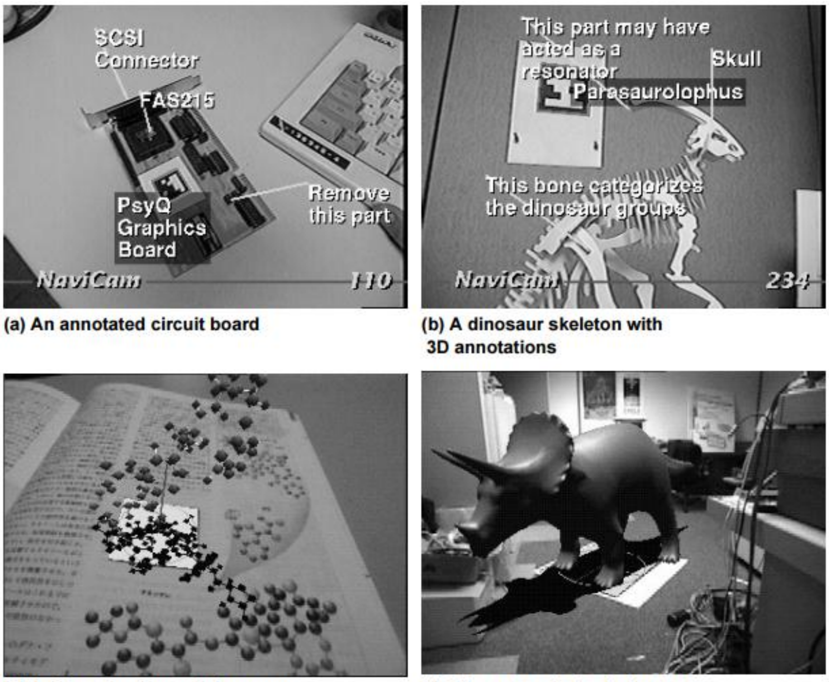
Jun Rekimoto et al.,[1] In augmented reality (AR) systems, it is crucial to correctly register real world computer information on the real world image. AR systems normally measure the position and orientation of a device with 3D sensors (either magnetic or ultrasonic); however these sensors often suffer from in an accuracy and limited tracking volume. Recently, vision based methods of estimating position information from known landmarks in the real world scene have been proposed. Bajura and Neumann used LEDs as landmarks and demonstrated vision-based registration for AR systems Aside that the position estimation problem, AR systems also need to determine which information should be overlaid without interfering with the user's real world tasks The 2D-matrix marker our method uses is a square shaped barcode that can identify 216 different objects. We can create and attach codes on a large number of real world objects at virtually no cost, because codes are printable with normal laser or ink-jet printers

In this paper introduces a new technique for producing augmented reality systems that simultaneously identify real world objects and estimate their coordinate systems. This method utilizes a 2D matrix marker, a square shaped barcode, which can identify a large number of objects. It also acts as a landmark to register information on real world images.

Finally they have presented a new method for realizing an augmented reality system by using printed 2D matrix codes. This method relays only on a video images, without requiring other 3D sensors.

A **Data Matrix** is a two-dimensional barcode consisting of black and white "cells" or modules arranged in either a square or rectangular pattern, also known as a matrix. The information to be encoded can be text or numeric data. Usual data size is from a few bytes up to 1556 bytes. The length of the encoded data depends on the number of cells in the matrix. Error correction codes are often used to increase reliability: even if one or more cells are damaged so it is unreadable, the message can still be read. A Data Matrix symbol can store up to 2,335 alphanumeric characters.

Data Matrix symbols are rectangular, usually square in shape and composed of square "cells" which represent bits. Depending on the coding used, a "light" cell represents a 0 and a "dark" cell is a 1, or vice versa. Every Data Matrix is composed of two solid adjacent borders in an "L" shape (called the "finder pattern") and two other borders consisting of alternating dark and light "cells" or modules (called the "timing pattern"). Within these borders are rows and columns of cells encoding information. The finder pattern is used to locate and orient the symbol while the timing pattern provides a count of the number of rows and columns in the symbol.



(a) An annotated circuit board

(b) A dinosaur skeleton with 3D annotations

(c) A 3D molecular model pops-up from the book

(d) Triceratops in the Laboratory (The matrix code is spreaded on the floor)

Fig 2.1 real-time object identification result

Daniel Wagner et al.,[2] Tracking from natural features is a complex problem and usually demands high computational power. It is therefore difficult to use natural feature tracking in mobile applications of Augmented Reality (AR), which must run with limited computational resources, such as on Tablet PCs. Mobile phones are very inexpensive, attractive targets for AR, but have even more limited performance than the aforementioned Tablet PCs. Phones are embedded systems with severe limitations in both the computational facilities (low throughput, no floating-point support) and memory bandwidth (limited storage, slow memory, tiny caches). Therefore, natural feature tracking on phones has largely been considered infeasible and has not been successfully demonstrated till date.

They have presented three techniques for 6DOF natural feature tracking in real time on mobile phones. They have achieved interactive frame rates of up to 30 Hz for natural feature tracking from textured planar targets on current generation phones. The have used an approach based on heavily modified state-of-the-art feature descriptors, namely SIFT and Ferns plus a template-matching-based tracker. While SIFT is known to be a strong, but computationally expensive feature descriptor, Ferns classification is fast, but requires large amounts of memory. This renders both original designs unsuitable for mobile phones. They have given detailed descriptions on how they have modified both approaches to make them suitable for mobile phones. The template-based tracker further increases the performance and robustness of the SIFT- and Ferns-based approaches.

They assumed that the simplicity of Phony Ferns would let it outperform the more complex Phony SIFT on a constrained platform such as a phone. However, it turned out that in order to deliver a high level of quality, Ferns requires significant amounts of memory (for a phone) and computational bandwidth to use the consumed memory. Moreover, the very simple structure of Ferns descriptors requires more sophisticated outlier management, which consumes further computational resources. The approach finally adopted for both shows interesting aspects of convergence: In both approaches, Laplacian Gaussian feature detection was replaced by simple FAST detector at the expense of losing scale independence. Phony Ferns adopted a regularization using the dominant orientation from SIFT, while Phony SIFT adopted a search forest approach from Ferns. Two of the three steps of outlier management, namely orientation check, and homographic check, are shared by both approaches. A major weakness of both approaches is the rather limited tilt angle (40-50degree) they can tolerate. This limit is strongly reduced by combining Phony SIFT and Phony Ferns with the Patch Tracker, which can still track at close to 90 degree tilt. A natural future step is to extend the presented work in order to support 3D tracking targets. In the case of 3D targets, estimating a homograph would not suffice anymore. Furthermore, it would be necessary to cope with self occlusions of the tracking target.

The **scale-invariant feature transform** (**SIFT**) is a feature detection algorithm in computer vision to detect and describe local features in images. SIFT keypoints of objects are first extracted from a set of reference images[2] and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of keypoints that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches. The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalised Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence

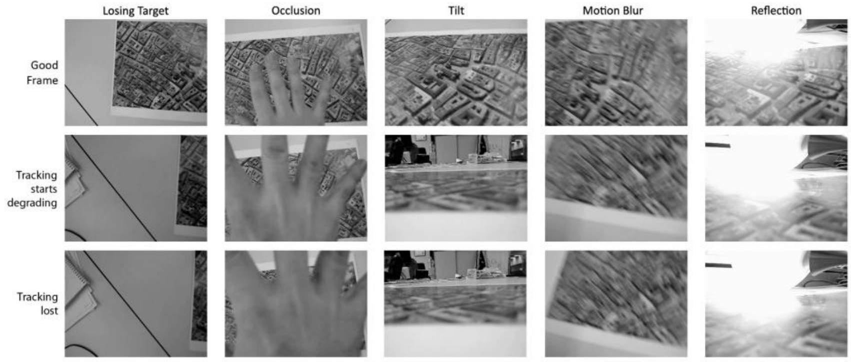


Fig 2.2 Real-Time Detection and Tracking

Hate Hagbi et at.,[3] Model-Based visual tracking has become increasingly attractive in recent years in many domains, such as robotics and Augmented Reality (AR). In many of these domains, visual tracking is often combined with object recognition tasks. In AR applications, model-based recognition and 3D pose estimation are often used for superposing computer-generated images over views of the real world in real time. Fiducial-based computer vision registration is popular in AR applications due to the simplicity and robustness it offers. Fiducials are of predefined shape, and commonly include a unique pattern for identification. Fiducials are useful for various tasks, such as prototyping and producing tangible interaction techniques for better user interfaces. On the other hand, Natural Feature Tracking (NFT) methods are becoming more common, as they are less obtrusive and do not require to modify the scene . This is achieved at the cost of increased computational complexity and decreased accuracy

In the paper has explained about their system called Nestor. Nestor is a real-time recognition and camera pose estimation system for planar shapes. The system allows shapes that carry contextual meanings for humans to be used as Augmented Reality (AR) tracking targets. The user can teach the system new shapes in real time. New shapes can be shown to the system frontally, or they can be automatically rectified according to previously learned shapes. Shapes can be automatically assigned virtual content by classification according to a shape class library. Nestor performs shape recognition by analyzing contour structures and generating projective-invariant signatures from their concavities. The concavities are further used to extract features for pose estimation and tracking. Pose refinement is carried out by minimizing the re-projection error between sample points on each image contour and its library counterpart. Sample points are matched by evolving an active contour in real time. Finally The system operates at interactive frame rates on a Nokia N95 mobile phone. It performs robust recognition of shapes and maintains accurate and stable 3D registration in extreme slant angles, as well as in the cases of partial occlusion. Nestor allows planar shapes to be used for registration as flexible Fiducials for AR. Nestor rectifies new shapes according to previously learned shapes and automatically assigns virtual content to them according to a shape class library.

ARToolKit is an open-source computer tracking library for creation of strong augmented reality applications that overlay virtual imagery on the real world in order to create strong augmented reality, it uses video tracking capabilities that calculate the real camera position and orientation relative to square physical markers or natural feature markers in real time. Once the real camera position is known a virtual camera can be positioned at the same point and 3D computer graphics models drawn exactly overlaid on the real marker. So ARToolKit solves two of the key problems in Augmented Reality; viewpoint tracking and virtual object interaction.

Fig 2.3 Shape Recognition

Taehee Lee et al.,[4] ,Augmented reality (AR) is a powerful human-computer interaction paradigm for wearable computing applications. The world around a mobile computer user can directly serve as the user interface, presenting a location-specific 3D interaction space where the user can display, examine, and manipulate information . A successful standard approach for viewing AR content and registering it with the world is via vision-based tracking of cardboard Fiducial markers ,which can be used as a hand-held tangible user interface for inspecting and manipulating the augmentations . Mobile AR research has produced many useful user interface options for wearable computing .

In the paper has presented marker less camera tracking and user interface methodology for readily inspecting augmented reality (AR) objects in wearable computing applications. Instead of a marker, they have used the human hand as a distinctive pattern that almost all wearable computer users have readily available. They have present a robust real-time algorithm that recognizes fingertips to reconstruct the six-degree-of-freedom camera pose relative to the user’s outstretched hand. A hand pose model is constructed in a one-time calibration step by measuring the fingertip positions in presence of ground-truth scale information. Through frame-by-frame reconstruction of the camera pose relative to the hand, They have stabilized 3D graphics annotations on top of the hand, allowing the user to inspect such virtual objects conveniently from different viewing angles in AR. They have evaluated their approach with regard to speed and accuracy, and compared it to state-of-the-art marker-based AR systems

They have segmented and tracked the hand region based on adaptively learned colour distributions. Accurate fingertip positions are then located on the contour of the hand by fitting ellipses around the segments of highest curvature. Their results show that camera pose estimation from the fingertip locations can be effectively used instead of marker-based tracking to implement a tangible UI for wearable computing and AR applications.

Markerless motion tracking: Emerging techniques and research in computer vision are leading to the rapid development of the markerless approach to motion capture. Markerless tracking do not require subjects to wear special equipment for tracking. Special computer algorithms are designed to allow the system to analyze multiple streams of optical input and identify human forms, breaking them down into constituent parts for tracking. It's possible to perform markerless tracking which continuously searches and compares the image with the known 3D model if the geometric characteristics of the target is known .

Fig 2. 4 marker less inspection result

Shaunak Shirish Deshmukh et al.,[5] Augmented Reality(AR) combines virtual world objects with real environments. Over years, augmented reality has been used in many domains for a multitude of purposes. In most fields, AR is used as an assistive system for performing human tasks. AR has proven to be useful in increasing the efficiency and accuracy of the tasks especially in the domains related to surgery and aeroplane manufacturing. In case of surgery, it can be used as a tool to render 3D models of the patients operated area / organ that can help doctors perform surgeries with minimum risk and complications. In aeroplane manufacturing, AR can be used as a tool to assist wiring the electrical harness of a plane which is a long and tedious task and is still done manually.

They have used AR to leverage the increased computing power of smart-phones to build a system that displays 3D objects using a printed image without using any complicated equipment. The purpose of this system is to accelerate learning and understanding of concepts such as structures or mechanisms. Instead of reading long manuals, the user can watch and interact with a 3D video manual through AR. The average person learns better by observing and listening something than by simply reading something. We will be using this specific property of the human mind to accelerate learning.

They have develop a system, that have used AR technology to improve learning and understanding of a product for electronics products consumers. In teaching, our system will help teachers to explain concepts to students effectively and enhance their learning. In the manufacturing sector this product will help workers understand a manufacturing process and assist them in doing so. Thier system aims at helping and developing better training systems, 3D user manuals and interactive courseware materials that will aid efficient learning and training for people

Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human vision system can do.

Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, *e.g.*, in the forms of decisions. Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that can interface with other thought processes and elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.

As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems.

D.W. F. van Krevelen et al.,[6] Augmented reality (AR) is this technology to create a next generation, reality-based interface" [78] and in factual ready exists, moving from laboratories around the world into various industries and consumer markets. AR supplements the real world with virtual (computer generated) objects that appear to coexist in the same space as the real world. AR was recognised by MIT as one of ten emerging technologies of 2007 , and with today's smart phones and AR browsers we are starting to embrace this very new and exciting kind of human-computer interaction.

In the paper, they have stated" We are on the verge of adopting Augmented Reality (AR) technologies to enhance our perception and help us see, hear, and feel our environments in new and enriched ways. AR will benefit us in fields such as education, maintenance, design, reconnaissance, to name but a few. We describe the field of AR, including a brief definition and development history, the enabling technologies and their characteristics. We survey the state of the art by reviewing some recent applications of AR technology as well as a number of known limitations regarding human factors in the use of AR systems that developers will need to overcome .Early mobile AR systems simply use mobile trackballs, track pads and gyroscopic mice to support continuous 2D pointing tasks. This is largely because the systems still use a WIMP interface and accurate gesturing to WIMP menus would otherwise require well tuned motor skills from the users. Ideally the number of extra devices that have to be carried around in mobile UIs is reduced, but this may be difficult with current mobile computing and UI technologies. They have surveyed the state of the art in AR technologies, applications and limitations. Specifically new to the AR research literature are the comparative table on displays (Table 1) and survey of frameworks and content authoring tools (Section 2.4). With over a hundred references this survey has become a comprehensive investigation into AR and hopefully provides a suitable starting point for readers new to the field. AR has come a long way but still has some distance to go before industries, the military and the general public will accept it as a familiar user interface. For example, Airbus CIMPA still struggles to get their AR systems for assembly support accepted by the workers . On the other hand, companies like Information in Place estimated that by 2014, 30% of mobile workers will be using augmented reality. Within 5-10 years, Feiner [58] believes that "augmented reality will have a more profound effect on the way in which we develop and interact with future computers."

They have surveyed the state of the art in AR technologies, applications and limitations. Specifically new to the AR research literature are the comparative table on displays and survey of frameworks and content authoring tools . With over a hundred references this survey has become a comprehensive investigation into AR and hopefully provides a suitable starting point for readers new to the field.

Fig 2.5 AR technologies result

Chen Guodong et al.,[7] Detecting and recognizing objects in images is one of the most difficult tasks in computer vision. Scale, rotation, viewpoint changes, occlusion and background clutter are the common challenges. Colour, intensity, gradient and depth are always used as the cues for object detection and recognition. Many objects can be accurately represented by their shapes.. In the past few years, contour based methods and shape feature based methods have been proposed. Most of the methods can be broadly classified as point-based approaches , boundary-curve based approaches or shape-region based approaches

Point based approaches is always limited to the noises and background clutter in the images. And it is hard to localize the region around the point as the spatial neighbourhood of the point. Point based approaches using examples always involving pure shape matching and handwritten digits . Shape region based methods have not been popular as features due to the sensitivity to segmentation errors.

They have proposed a novel framework for object detection. In the training stage, the object class model can be represented by the codebook of shape fragments. By comparing the shape fragments of the detect image to the codebook, a distribution of object centre image is successfully achieved. Experimental comparisons with other methods were carried out to evaluate the proposed method, and test results shows the method are more distinctive and robust to image transformation and background clutter.

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

Fig 2.6 Model based Object Detection

Omer Akgul et al.,[9] As a man-machine interface technology, AR (Augmented Reality) simply renders virtual information on to real objects [1]. In order to achieve geometrically valid rendering, one needs to track the object of interest to be augmented. This can in principle be done using computer vision techniques. The target object of interest can be detected and tracked in a live video stream. The target can be a simple planar marker or any three-dimensional (3D) object . Known model of the object can be used to determine the position and orientation of the object. Rendering of the virtual object follows easily.

In practice, two-dimensional (2D) planar targets are frequently used as an object of interest . These type of targets are easy to track. Many algorithms have been proposed that detect the target in the given image and track it in the consecutive images. Detection algorithms employ a feature-based approach where interest points are matched with that of the reference views .Recent advances in hardware and algorithms have sparked an interest in deep learning algorithms. Deep learning methods have successfully been used in complex recognition tasks such as written digits recognition and object classification.

A known target, in this case a planar object, is rendered under various viewing conditions including varying orientation, scale, illumination and sensor noise. The resulting corpus is used to train a convolution neural network to match given patches in an incoming image. The results show comparable or better performance compared to state of art methods. Timing performance of the detector needs improvement but when considered in conjunction with the robust pose estimation process promising results are shown.

The target image is rendered to create many synthetic views from different angles and under different illumination conditions. The detection performance is shown to be comparable (if not better than) to one of the best algorithms in the literature. The detector performance is very strong especially in the presence of error in feature localization. The method can also be tailored to applications when the viewing geometry and illumination conditions are known in advance. In this case, the rendered training views can be customized to reflect the needs of the application.

In deep learning, a **convolution neural network** (**CNN**, or **ConvNet**) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery.

CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as **shift invariant** or **space invariant artificial neural networks** (**SIANN**), based on their shared-weights architecture and translation invariance characteristics

Convolution networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage.

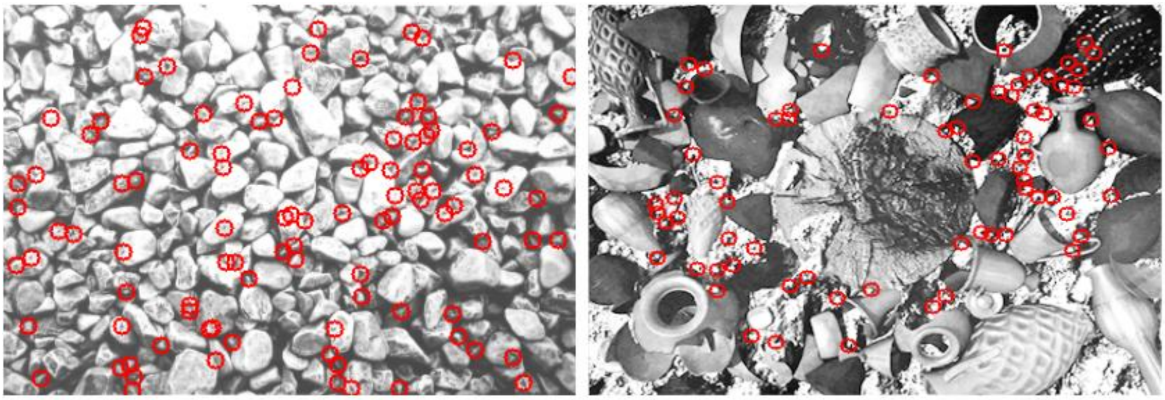


Fig 2.7 Deep Learning in Augmented Reality result

ShengzhongFeng et al.,[9] Pervasive games are computer games that have one or more salient features that expand the traditional spatial boundaries of games, temporally or socially . They represent an exciting development in playing games which leverages the use of sensor; visualization and networking technologies provide immerse live-action game experiences , and extend the game experience out into the real world . Nowadays, handheld and wearable devices supported by advance in multiple sensing systems, have given rise to pervasive computing in which multimedia information is embedded into the physical world around us . Meanwhile, natural interaction such as gestures, body movement, gaze and physical awareness are increasingly required and they are started to be applied, yet their true potential for intuitive use is little known by people . Genres of pervasive games are admittedly classified as augmented/mixed reality, pure location-based, mobile and trans-reality games They have presented their opinion that there is an increasing interest in creating pervasive games based on emerging interaction technologies. In order to develop touch-less, interactive and augmented reality games on vision-based wearable device, a touch-less motion interaction technology is designed and evaluated in this work. Users interact with the augmented reality games with dynamic hands/feet gestures in front of the camera, which triggers the interaction event to interact with the virtual object in the scene. Three primitive augmented reality games with eleven dynamic gestures are developed based on the proposed touch-less interaction technology as proof. At last, a comparing evaluation is proposed to demonstrate the social acceptability and usability of the touch-less approach, running on a hybrid wearable framework or with Google Glass, as well as workload assessment, user's emotions and satisfaction

A series of designed gestures and the evaluation research related to touch-less interaction technology on vision-based wearable device are proposed in this paper. The proposed touch-less interaction technology offers a full view of body control mode on wearable devices and that is the difference from traditional interaction technology. By comparing result of user studies, it can be revealed that the smart glass is a preferable platform for touch-less interaction approach. Although they are fun to use present touch-less games based on augmented reality on wearable devices have certain limitation. For example, their performance are limited to hands and feet gestures. Normal full-body gestures are even not considered in the research version.

Google glass was an image recognition mobile app developed by Google. It was used for searches based on pictures taken by handheld devices. For example, taking a picture of a famous landmark searches for information about it, or taking a picture of a product's barcode would search for information on the product .The system could identify various labels or landmarks, allowing users to learn about such items without needing a text-based search. The system could identify products barcodes or labels that allow users to search for similar products and prices, and save codes for future reference, similar to the Cue Cat from late 1990s. The system also recognized printed text and uses optical character recognition (OCR) to produce a text snippet, and in some cases even translate the snippet into another language.

Fig 2.8 Touch-less interactive augmented reality

Kristensen et al.,[10] Children’s gaming habits have changed significantly during the past decades (Zagal et al., 2000), from board games over computer games to mobile phone gaming and other types of games involving pervasive computing. The goal of the Battle Board 3D project is to combine the “best” features from classical board games and computer games. It’s their thesis that by augmenting a board game the aesthetic game experience will be extended. In a report on children’s use of computer games points out that a great amount of leisure time is used on these. In the fulfilment of this vision they turned to the technologies and perspectives of Augmented Reality. The social interaction has likewise changed from an intimate game experience in board games to a solitary game experience in the gaming arcades and behind the screens at home, and now multiplayer games over the Internet, multiplayer mobile phone games etc.

They have discussed the design of Battle board 3D (BB3D) which is an AR Toolkit based game prototype, featuring the use of LEGO bricks for the physical and digital pieces. BB3D is a novel type of an AR game augmenting traditional board games with features from computer games. The initial experiments involving kids indicate that it is promising with respect to add computer games excitement to board games and to add a social dimension to computer games. The paper discusses the concept for the game, implementation issues, the physical setting for the game, user interfaces, as well as tailor able pieces and warriors. Based on qualitative experiments with children, we discuss central design issues for future AR board games.

Their experiments show that the children found the combinations of some of the social interaction around classical board games and dynamic computer games exiting and amusing. Their design of the physical pieces has both some weak and some powerful features. The patterns on the pieces are designed to be human readable, and the pieces are designed to be picked up and moved freely on the board. However, the need for breaking up the pieces to trigger a battle breaks the game flow, but is also a suspense giving element. Further research in the area of how to trigger battles is necessary to find out which method is the most suitable. The pieces in Battle Board 3D are larger than in ordinary board games, a wish for smaller pieces was put forward by the children. This aspect should be developed further. There are also a number of computer game features that could be integrated in BB3D, like levels, single player mode, and energy state for the digital figures.

ARToolKit is an open-source computer tracking library for creation of strong augmented reality applications that overlay virtual imagery on the real world. n order to create strong augmented reality, it uses video tracking capabilities that calculate the real camera position and orientation relative to square physical markers or natural feature markers in real time. Once the real camera position is known a virtual camera can be positioned at the same point and 3D computer graphics models drawn exactly overlaid on the real marker. So ARToolKit solves two of the key problems in Augmented Reality; viewpoint tracking and virtual object interaction.

RESULTS:

Fig 2.9 Designing an Augmented Reality Board Games with children

CHAPTER 3

SOFTWARE REQUIREMENT AND SPECIFICATION

3.1 Hardware Requirements

Processor	Intel Core i5 or AMD FX 8 core series with clock speed of 2.4 GHz or above
RAM	2GB or above
Hard disk	40 GB or above
Input device	Keyboard or mouse or compatible pointing devices
Display	XGA (1024*768 pixels) or higher resolution monitor with 32 bit color settings
Miscellaneous	USB Interface, Power adapter, etc

Table 3.1: Hardware Requirements

3.2 Software Requirements

Operating System	Windows
Programming Language – Backend	Python , open CV module
Development environment	Spyder 3.6

Table 3.2: Software Requirements

CHAPTER 4

PROBLEM STATEMENT

Our project topic is "object recognition and visual innovation using augmented reality". We focus on various domains like augmented reality systems, object tracking, recognition or detection systems.

Augmented reality (AR) can turn everyday objects, places, and images, into new, interactive opportunities to engage with their customers using attractive graphics, video, animation, audio, and 3D content that brings in real results.

Many industries have adopted AR already and now it is gradually penetrating in marketing and sales with its power to revolutionise. So we decided to do this project on augmented reality which also involves detecting and recognising real world objects using deep learning tools like tensor flow, convolution neural network with AR toolkits like vuforia SDK.

In augmented reality applications, tracking and registration of both cameras and objects is required because, to combine real and rendered scenes, we must project synthetic models at the right location in real images. Although much work has been done to track objects of interest, initialization of theses trackers often remains manual. Our work aims at automating this step by integrating object recognition and other visual innovation in an AR system.

The idea behind it is very similar to Image Recognition, but instead of recognizing images and planar surfaces, this feature can work with three-dimensional structures and objects. We focus on detecting In particular, the initialization of the object tracking often remains manual in most systems.

We have planned on using the Bayesian Networks to perform the initialization phase of the tracking. By recognising the object, special points are taken and we use this information to create generic markers around the scene. Then, an algorithm for pose estimation is used to find the orientation of the real object to allow the registration process for 3D objects. The tracking and registration of both cameras and objects are fundamental tasks specially because some applications may be outdoor. and recognition real world objects and to innovative using various AR techniques.

This project will help us build an augmented reality system which can detect, recognise real world objects and can implement AR facilitated navigation, visual innovation and other information system.

CHAPTER 5

METHODOLOGY

In this project, we propose a simple and intuitive approach to AR content creation that comprises two steps: creation of 3D models from real objects using an image-based modeling tool (Autodesk Memento) and conversion of the 3D model (Aumentaty Author), into an interactive AR element using an AR authoring tool.

Creation of 3D Models

The first step requires the acquisition of 3D information of the object that needs to be modeled. Because of the nature of image-based modeling software, objects with plain, transparent, glossy, or reflective surfaces will not work correctly. Multiple pictures will be taken by shooting at least a loop of sequential photographs about the subject. The set of photographs can now be processed by an image based modeling tool. In this project, we used Autodesk Memento Beta, a solution for converting captured reality input into high definition meshes that can be fixed and optimized. Depending on the quality of the resulting 3D model, additional cleanup may be required to eliminate unnecessary noise or busy surroundings. Basic fixing/cleanup can be performed by smart selection and clean up tools available in Autodesk Memento. Finally, the textured 3D model reconstructed by Autodesk Memento can be exported to several formats: OBJ, STL, PLY, and FBX. Because of the formats supported by our AR authoring tool and based on Aumentaty recommendations, models are exported to FBX so they can be processed successfully.

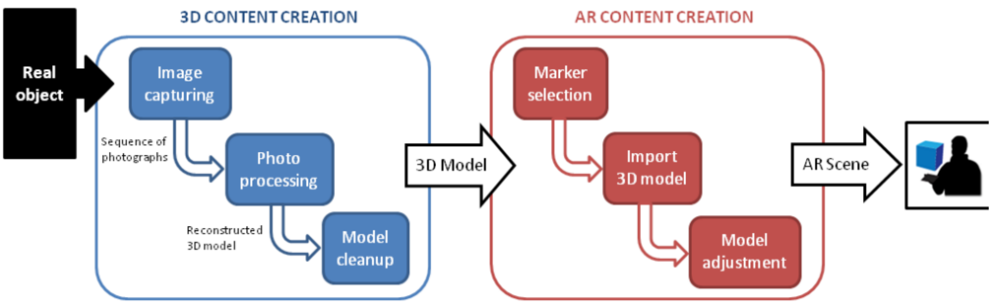


Fig 5.1 Flow diagram of Project Working

Creation of AR Content

Creating an interactive AR element with Aumentaty Authoris a visual and intuitive process. First, a marker is selected from the marker ID menu and printed, so the scene can be visualized interactively as it is being created. Next, the camera needs to be activated and pointed to the printed marker. On the computer screen, the marker will turn orange, indicating that the AR software is recognizing the marker. At this point, the user can browse for the model that was created in the previous step and drag and drop it from the models library part of the interface to the icon that corresponds with the marker that was printed in the first step. The 3D model will be placed over the marker in the software interface to provide a visual cue of the link that was created

The controllers located on the main panel can be used to move, scale, and rotate the 3D with respect to the marker. This allows for basic adjustments in the scene, which are useful in many situations such as when the model has a different vertical orientation from the marker's scene or the 3D model is too large with respect to the marker.

Finally, the AR scene with all markers and models will be exported to Aumentaty Viewer, so it can be visualized on any device (desktop or mobile) equipped with a camera

CHAPTER 6

APPLICATION

1. Empower the Gaming Industry:

The gaming industry is the biggest beneficiary of augmented reality. We have already seen Pokémon Go as a remarkable example of using augmented reality in gaming. Mobile gaming is a growing market and augmented reality has added a scope for high-end games to engage mobile users.

Fig 6.1 Pokémon go

2. E-commerce will be redefined:

Up until now, we could only see a product image online and that was supposed to be the basis for making purchase decisions. With augmented reality, we can now combine the convenience of online shopping with the reliability of purchasing from physical stores. AR tools have enabled people to try new outfits without going through the hassle of changing clothes. IKEA is also a good example of enabling people to understand how a furniture will look at home thus, enabling people to make better purchase decisions.

Fig 6.2 E-commerce using AR

3. Power o utility apps:

While utility apps haven't made significant advancements in AR, there are immense potential ways of using it. Google translate, for example, allows you to use your camera to translate text on billboards and street signs. Augmented reality can also be used for navigation when combined with helmets. This will allow the wearer to see a pointer indicating direction, eliminating the need to look at the mobile phone.

Fig 6.3 AR utility app1

\

Fig 6.4 AR utility app 2

CHAPTER 7

FUTURE ENHANCEMENTS AND SCOPE

An area that might be revolutionized by AR is the retail sector, where the technology will breach the gap between online and physical shops, offering a richer experience and personalized promotions delivered right to the wearer. Computer games are obviously another important area of development for AR.

The mobile industry is also a potential scene for the future of AR. From smart phones accessing GPS to wearable technologies ranging from Google Glass to Apple Watch, augmented reality will be expressed by specific apps and micro-location technologies like iBeacons. All will eventually blur the distinction between reality and virtuality – bringing users a host of relevant data to completely alter their experience. There are already apps which offer some basic versions of this functionality on Smartphone.

Some other possible future uses of augmented reality are in the sector of medicine (direct training of students), military (precise location of enemy positions), fine art (mixed media and interactive artworks), industry (assistance in prototyping) and teaching (visualizing learning material in interactive apps).

CHAPTER 8

CONCLUSION

In first phase of our project work we learnt how augmented reality can be implement for our project application by doing literature survey using existing base papers. In further days by having all these ideas about our application we will start implementing it and try to get the best outcome.

When a person is curious, he searches all the contents about the object is he curious on. We in this project are making it simpler and easier for the person using it, in the world of augmented reality he gets the extra information automatically other than trying various ways to search it. After completion our project works will work in below explained manner.

If a person wants to find out some detail about an object of some kind , suppose the person wants to know the details of a museum, so he opens his camera to detect the place and clicks a picture of that museum ,the phone uses its built-in global positioning system(GPS) to figure out roughly where the person is ,then quickly searches Google images to find similar photos taken in the same neighbourhood. In couple of seconds, its identified the building and brought up a Wikipedia page telling all the details.

Another case is getting the details of an object. The person points his phone at the object and has some kind of pattern recognition or feature detection system to identify it. Once the object is identified the background system gets into work of getting all the details of that object. In a couple of seconds the user will see the details of that object next to it.

REFERENCES

[1]. Jun Rekimoto, **"Matrix: a realtime object identification and registration method for augmented reality",** IEEE, August 2002 .

[2]. Daniel Wagner," **Real-Time Detection and Tracking for Augmented Reality on Mobile Phones** " ,IEEE transactions on visualization and computer graphics, vol. 16, no. 3, may/june 2010.

[3]. Nate Hagbi, **"Shape Recognition and Pose Estimation for Mobile Augmented Reality",** IEEE transactions on visualization and computer graphics, vol. 17, no. 10, october 2011.

[4]. Taehee Lee, Tobias Hollerer ," **Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking**" ,IEEE 2007.

[5]. Shaunak Shirish Deshmukh, Chinmay Mandar Joshi, Rafiuddin Salim Patel , Dr. Y. B. Gurav,**" 3D Object Tracking And Manipulation In Augmented Reality",** International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue

[6]. D.W.F.van Krevelen,R.poelman, "**A Survey of Augmented Relity technologies Applications and Limitations",** june 2010.

[7] Chen Guodong, ZeyangXia, Rongchuan Sun, Zhenhua Wang, Zhiwu Ren and Lining Sun, "**A Learning Algorithm for Model based Object Detection",** 2011 IEEE.

[8] Omer Akgul, H. Ibrahim Penekli and Yakup Genc, **"Applying Deep Learning in Augmented Reality Tracking"** , IEEE International Conference on Signal-Image Technology & Internet-Based Systems, 2016

[9] Zhihan Lv, Alaa Halawani, Shengzhong Feng, Shafig Ur , Haibo Li , "**Touch-less interactive augmented reality game on vision-based wearable device**", Journal Personal and Ubiquitous Computing Volume 19 Issue 3-4, July 2015.

[10] Grobaek,kaj,Kristensen,sune,Andersen,Troels L. Nielsen,Bjorn W, **"Designing an Augmented Reality Board Games with children: The BattleBoard 3D experience"**Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community, 2004.