**­­­Augmented Reality Rubik Cube Solve Design Doc**

[1 - Introduction 1](#_Toc405303409)

[2 - Goals 1](#_Toc405303410)

[3 - Architecture 2](#_Toc405303411)

[4 - Design 2](#_Toc405303412)

[5 - Event Model 3](#_Toc405303413)

[6 - Architecture 4](#_Toc405303414)

[7 - Image Analysis Design 4](#_Toc405303415)

[8 - Class Diagram of Persistent Objects 5](#_Toc405303416)

[9 - Class Diagram of Per-Frame Objects 6](#_Toc405303417)

[10 - Parameters 6](#_Toc405303418)

[11 - Notes 7](#_Toc405303419)

[12 - Stable Face Recognizer State Diagram 7](#_Toc405303420)

[13 - Application State Machine State Diagram 8](#_Toc405303421)

# Introduction

This document describes the associated Android application that can visually recognize a Rubik Cube and guide the user through edge rotation steps to return a cube to its’ original configuration. The application is written nearly completely in Java and makes heavy use of OpenCV for image recognition. This document describes the architecture, design, implementation and performance achievements of this application.

# Goals

## Solve Rubik Cube from Arbitrary Starting Position

This is of course the main purpose of the application.

## Target Platform

While developed on a standard commercial Android Smart Phone, the target platform is a set of Fully Immersive Smart Glasses such as CastAR from Technical Illusions, Moverio BT-200 from Epson, Meta from Space Glasses, and ATHEER AiR™ from Atheer One to name a few.

## Operation in a natural environment

The application must work in a normal cluttered environment without necessity of specific background or specific lighting. However, with that said, there will exist some backgrounds and some lighting conditions under which the application will not perform. It is the long term goal of this project to operate as robustly as possible in as many different environmental conditions as possible.

## Gesture Driven

In normal mode there are no menus or text displays, rather the application recognizes the current position and status of the cube and provides graphics instructions on how to solve the cube.

## Productization

It is undecided at the time of this writing whether this application will be published on Google Play Store for general use. It does not make sense to run this application on a Smart Phone or Tablet since two hands are required to rotate the cube and an additional hand to hold the device. Smart Glasses, especially Android centric devices, are in their infancy and little standardization has yet occurred. In all probability, Augmented Reality application like this will need to be hand ported to each and every Smart Glasses device for some time to come.

# Architecture

In reality, the core architecture is just the simple “Model-View-Controller” pattern, or for this application, more logically phrased as a “Controller-Model-View” pattern. Specifically:

## Controller

This lion-share of the application complexity in both code and CPU consumption is in this domain. Specifically, each incoming Camera Image Frame is fully analyzed in an attempt to recognize a Rubik Face. Then, additional processing is performed in class StateMachine to determine if the Model (or State) should be changed, and, when all six sides of the Rubik Cube have been recorded, the TwoPhase[] Rubik Cube Logic solution is called.

## Model (or State)

Rubik Cube state is captured in the simple class StateModel.java. Besides the cube six faces, a small amount of additional data (state) is held in this class. Otherwise, this represents pretty much the full state of the system.

## View

There are actually four separate “Views,” they are:

### Diagnostic Annotations

A wide range of diagnostic information is available through the Menu->Annotation user interface. This is rendered using OpenCV on the frame image itself. Information broader that what is in state is offered.

### Pilot Cube

At the time of this writing, the Pilot Cube is immature, but its long term purpose is to represent the state of the cube both in tiles and faces, and eventually to show requested Cube Rotations. This is rendered using OpenGL.

### Requested Text Directions

Text instructions to user to rotate entire cube or rotate an edge can be optionally displayed as a text message. This is rendered using OpenCV on the frame image itself.

### Requested Arrow Directions

Graphic instructions in the form of 3D arrows are rendered in a fashion overlaying the cube itself with proper position, orientation and size.

# Design

Clearly the “Controller” is the more interesting and challenging aspect of this application.

## Image Processing

Also correlates to what OpenCV offers.

### Grey Scale

### Gaussian Blur

### Edge Detection

### Contour Detection

### Polygon Detection

## Face Recognition

### Polygon Detection

### 3x3 Tile Grid Reconstruction

### Tile Color Detection

## 3D Reconstruction

Purpose is to produce OpenGL required cube translation and rotation.

### Definition of Cube in GL space

### Homography and Perspective Transformation equations

### Over constrained Non-Linear equations with six unknowns

## Gesture Recognition

### Stable Face Detection

### Stable New Face Detection

## Application State Machine

### Guided Exploration Phase

### Solution Computation

### Guided Solution Phase

# Event Model

## Thread Model

Main Thread: - That which constructs the app?

UI Thread: - Menu events are handled here

OpenCV Frame Thread: - Camera Frame image is provided and processed.

OpenGL Thread: - Separate thread on which OpenGL renderers are executed.

## Tracking and Feedback

For simplicity, no frame to frame information path exists. That is, each frame is processed without any information from the previous frame.

# Class Diagram of Persistent Objects

**+**

Activity

Menu and Params

Image Recognizer

App State Machine

State Model

Face

Instructions View

Overlay View

Pilot Cube View

# Class Diagram of Per-Frame Objects

**+**

**+**

**+**

Rubik Face

Rhombus

Polygon

Contour

Constant Tile Color

Least Means Square

Profiler

Frame

Measured Tile Color

# Parameters

The image recognition process uses many “offline parameters” that can be adjusted to improve the process. Manual adjustment is, of course, undesired. However, achieving automatic (i.e., “online”) determination of these parameters shall be part of the maturing of this application.

All determined, more or less, empirically.

These parameters can be adjusted using the menu option “parameters.” All parameters are modeled as double precision floating point, and the purpose they serve in the application is as follows:

## Gaussian Blur Kernel Size

The square kernel size of the Gaussian Filter. The default is set to 3. The filter sigma will be computed from the kernel size using the formula: sigma = 0.3\*((ksize-1)\*0.5 - 1) + 0.8.

## Canny Edge Upper Threshold

Gradient threshold at which an edge is declared to exist.

## Canny Edge Lower Threshold

Gradient threshold at which an edge (once detected) is declared to no longer exist.

## Dilation Kernel Size

## Minimum Contour Area Size

Minimum contour area size to be considered a tile candidate.

## Polygon Epsilon Threshold

## Minimum Parallelogram Area Size

Minimum parallelogram area size to be considered a tile candidate.

## Maximum Parallelogram Area Size

Maximum parallelogram area size to be considered a tile candidate.

## Parallelogram Angle Outlier Threshold

The angles (in degrees, there are two angles) that the parallelogram sides make with respect to the absolute coordinate system is calculated. The mean of these angles among the entire candidate set is calculated. They should roughly have all the same value. If any parallelogram exhibits a sufficiently different shape (i.e., angle of sizes) it is rejected.

## Face Least Means Square Sigma Threshold

Threshold for residual Least Means Square error of fitting Parallelogram centers into a 3x3 grid as defined by their candidate grid location and their actual measured centers. If the residual is less than this value, the candidate tile grid locations are considered valid.

## Manual Luminous Offset

# Notes

* Prefix Rubik on all allows for easy identification and use of common names like Util and Controller.
* Significant use of public key word on member variables.
* Public vs Private: Member data is often made public in this implementation. Compliance with the convention of keeping member data private as much as possible results in too much code being in a very few files: particularly cube state and face state information. The convention of getters and setters is possible, but really accomplished nothing in terms of member data containment. I believe this issue is practically an outcome of the use of the MVC design pattern.
* Inter UI vs Frame thread contention handled by use of simple Boolean variables: i.e., atomic by nature.

# Stable Face Recognizer State Diagram

Other State Member Data:

Last :== Last Valid Face Object

Count :== Integer for counting

On Face Event

[ Face == null ]

Count = 0

[ Face != null ]

Last = Face

Count ++

[ Face == Last && Count < 4 ]

Count ++

[ Face == Last && Count >= 4 ]

[ Face == Last ]

[ Face != Last ]

[ Face != Last]

Count = 0

[ Face == Last ]

[ Face != Last && Count < 4 ]

Count ++

[ Face != Last && Count >= 4 ]

On Stable Face Event

Off Stable Face Event

# Application State Machine State Diagram

Other State Member Data:

Status :== True if all six sides of Rubik Cube have been observed

Count :== Integer for counting

On Face Event

Off Stable Face Event

On Stable Face Event

Count= 0

[ Count >= 4 ]

[ Count < 4 ]

Count++

[ Status == False ]

Count = 0

[ Status == True && colors OK ]

[ Status == True && colors ! OK ]

## App State Definitions

START: Ready

GOT\_IT: A Cube Face has been recognized and captured.

ROTATE: Request user to rotate Cube.

SEARCHING: Attempting to lock onto new Cube Face.

COMPLETE: All six faces have been captured, and we seem to have valid color.

BAD\_COLORS: All six faces have been captured, but we do not have properly nine tiles of each color.

VERIFIED: Two Phase solution has verified that the cube tile/colors/positions are a valid cube.

WAITING: Waiting for TwoPhase Prune Tree generation to complete.

INCORRECT: Two Phase solution has analyzed the cube and found it to be invalid.

SOLVED: Two Phase solution has analyzed the cube and found a solution.

DO\_MOVE: Inform user to perform a face rotation

WAITING\_FOR\_MOVE\_COMPLETE: Wait for face rotation to complete

DONE Cube should be completely physically solved.

# 2D to 3D Reconstruction

Since the shape of the Rubik Cube is well defined, from simply one face (always the up face) tile location in 2D space, the position and orientation of the cube in 3D space can be calculated.

## Rubik Cube definition in Real World Coordinates

By convention, the top face is the face that is recognized by the system.

**X**

**Y**

**Z**

The four corners of the top face are defined, in OpenGL Space, to be at locations:

{+1.0, +1.0, +1.0}

{-1.0, +1.0, +1.0}

{+1.0, +1.0, -1.0}

{-1.0, +1.0, -1.0}

The center of the nine tiles on the top face a defined, in OpenGL Space, to be at locations:

{+0.666, +1.0, +0.666}

{ 0.000, +1.0, +0.666}

{-0.666, +1.0, +0.666}

{+0.666, +1.0, +0.000}

{ 0.000, +1.0, +0.000}

{-0.666, +1.0, +0.000}

{+0.666, +1.0, -0.666}

{ 0.000, +1.0, -0.666}

{-0.666, +1.0, -0.666}

**X**

**Y**

**M**

**N**

Image

Space

Coordinates

Face

Indicie

Coordinates