School of Computing

Year 4 Project Proposal Form

SECTION A

Project Title: Stolen Image Identification

Student Name: Thomas Doyle

Student ID: 15350316 Stream: CASE

Project Supervisor Name: Brian Stone

SECTION B

General area covered by the project:

This project will cover the area of computer vision and image classification. It will be primarily to research known methods of comparing images and making a prediction to whether one is a copy of the other or it has elements taken from another image and placed in a new one. Testing will be done on numerous algorithms to come up with a final solution that is fast but also correct.

Outline of the proposed project:

This project will attempt to identify how similar two images are to determine the likelihood that one is a stolen version of the other. This will take an image supplied by the user and test it against a database of known metadata about images. It will then give an opinion if these images are sufficiently different one is in fact a copy of the other. This will be done in the form of a percentage of confidence determined by an algorithm.

Background:

Photographers always have a fear when publishing images that they might be taken and used without permission, watermarks edited out, or parts of images taken to create new images. This results in a loss of revenue for this photographer. Images are easy to steal and hard to track down, especially the source of an image. With this automated search tool it will take out the human element of comparing and contrasting the image to attempt to find duplicate copies.

Achievements

This should be collecting data about images constantly and storing the metadata and computed data about the image. This will then generate a report of stolen images and alert the origin of these images.
It will also allow a user to find the origin of an image if the project has seen it before.

Justification

This will allow photographer to search if anywhere else is using their images. It will also allow me to research the area of known image comparison algorithms and attempt to implement and verify this work.

Programming language(s)

The programming language I will use is python. I feel it has many libraries for production ready computer vision and it is very fast for me personally to write and develop in, which may be necessary because i will be implementing far more than the final algorithm to test which one suits the situation best. Given the short time frame of the project I don't believe I have the time to learn and become comfortable in a new language.

Programming tools / Tech stack

For my database I am going to choose postgreSQL, apache web server, imageAI library for python computer vision and most likely make use of Amazon Web Services for the high scalability.

Learning Challenges
I will have to learn how to design the entire project to scale with a vast amount of data
I will have to learn computer vision and comparison algorithms
I will have to learn how to apply machine learning to continuously improve the algorithms

Hardware / software platform
The software will be design to run on linux on cloud hosted machines
There is no special hardware for this project