

# GOOGLE™ SEARCH BY IMAGE: A SYSTEM EVALUATION OF ADJUSTED IMAGES FOR THE DETECTION OF VISUAL PLAGIARISM

L. VAN HEERDEN, S.I. DUMINY & N.J. LUWES

CENTRAL UNIVERSITY OF TECHNOLOGY, FREE STATE, SOUTH AFRICA

## Abstract

This paper investigates the precision of Google™'s Search by Image (SBI) system which lecturers can use to establish a workflow that will combat visual plagiarism in photography programmes. Currently no efficacious visual plagiarism detection method exists for implementation by photography lecturers. Content-based image retrieval systems like Google™ SBI have not yet been tested systemically for the detection of visual plagiarism. Using the Precision method to calculate the accuracy of the system, 300 images were randomly sampled through Google™ Images and altered with different adjustments. The images were uploaded to Google™ SBI and the results indicated a system of high quality.

**Keywords:** photography education, image plagiarism, visual plagiarism detection, Google™ Search by Image, content-based image retrieval

## 1. INTRODUCTION

Visual plagiarism in photography programmes is an issue that discriminates neither demographically nor geographically. The Internet provides unrestricted access to media to every person wishing to exploit it, causing a rise in visual and text-based plagiarism (Gutenko 2000, Garrett & Robinson 2012, Stoltenkamp & Kabaka 2014). Brian Kates, cited by Howard (2007: 3), voiced this concern with the rise in Internet usage among students:

*“In numbers growing by the thousands, students have found a quick-fix cure for their academic headaches — on the Internet. In the wonderful world of Web sites, scores of online companies are eager and able to provide slackers with whatever they need — for a price.”*

Garrett and Robinson (2012: 2) found that not only did Internet usage cause a rise in visual plagiarism in visual arts programmes like photography, but also that institutions were failing to detect and deter it. They developed a visual plagiarism identification pilot called iTrace, which performed well during testing on overall functionality and service, but the participants were concerned with its usefulness to a photography lecturer, as the service was not extended across the whole Internet and limited to a database selected by the user.

Canvassing a specific corpus of images, for instance of previous students' works, may not be effective, as the study found that most students plagiarise by copying images from Internet sources and presenting those images as their own. In some cases students would digitally alter the images with basic adjustments in an attempt to obfuscate the origin. The most common of these alterations found by Garrett and Robinson (2012: 8) were colour and contrast adjustments.

Currently, the only weapon lecturers possess against visual plagiarism is becoming familiar with students' performance limitations and styles that suspect images may be addressed. Even if the lecturer suspects visual plagiarism, proving it is difficult, as very little research has been undertaken in the detection of visual plagiarism compared to text-based plagiarism (Garrett & Robinson, 2012: 1).

This paper will suggest using Google™ Search by Image (SBI) as a visual plagiarism detection tool to assist photography lecturers in identifying plagiarised digital images by means of a system evaluation. Google™ SBI will match the suspect uploaded image to its possible source on the Internet. The purpose here is to test the success rate of the Google™ SBI system as an effective visual plagiarism detection tool and provide photography lecturers with a feasible method of identifying plagiarised images from any location where an Internet connection is available. In order to retrieve digital images from its very large database, Google™ SBI makes use of a computer vision technique called content-based image retrieval (CBIR).

CBIR is any technology that organizes digital image archives by their visual content (Datta, Joshi, Li & Wang, 2008: 2). A CBIR system functioning at a basic level extracts low-level features from the query image uploaded by the user and assigns a weight to each feature. It then combines the weights as per its design and retrieves the nearest neighbouring images as per the rank (Vadivel, Majumdar & Sural, 2004: 128). The low-level features include colour, shape and texture. A CBIR system functioning at a higher level may also extract salient points or be able to identify the objects in the image by some degree of logical inference similar to the way new digital cameras recognize facial features. Google™ SBI functions at a higher level, as it employs interest point detectors along with low-level feature detection.

Google™ SBI also combines content-based and textual image retrieval. This fusion of techniques helps CBIR systems that have to canvas very massive databases by limiting the pool of images first through metadata and then further narrowing the results by visual content. This may not be the ideal solution, as a lot of images on the Internet may be tagged incorrectly or subjectively, and these images will remain hidden from the retrieval process. Automated annotation is one part of solving the problem, but there still remains no means of creating a universally acceptable algorithm that characterizes human vision.

This problem is referred to as the semantic gap. According to Datta et al. (2008: 14), this is “the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation”. Rui, Huang, Ortega and Mehrotra (1998: 1) describe the semantic gap as “the gap between high-level concepts and low-level features”. In order to create a completely successful CBIR system that can be used in the real world, research in this area needs to focus on bridging the semantic gap by finding an absolute way of describing an image mathematically and assessing the similarities between two visual images based on their abstract descriptions. Datta et al. (2008: 24) suggest that reducing this gap should be the main research goal in CBIR technology's future. This can only be made possible if the research communities who will use CBIR systems test them for real-world applications so that the designers can have better feedback of what is needed.

The aim of this study is to determine Google™ SBI's accuracy in detecting images that were altered with basic adjustments. The resultant data will inform the photography lecturer of the trustworthiness of the system's ability to detect images that were unethically appropriated by students from the Internet, even if those images were altered from the original with the basic alterations found by Garrett and Robinson (2012: 8). These basic alterations are colour and contrast adjustments. So, the following three alterations were applied to determine to what degree of accuracy the CBIR system, Google™ SBI, can retrieve such images successfully:

- Black and white conversion (The absence of colour)
- Contrast increase
- Hue shift (A shift in colour)

Following is a breakdown of the research methodology used to evaluate the Google™ SBI system.

## **2. METHOD**

### **2.1 Research Design**

An adapted approach of Simulating Users was applied to obtain the data, which, in turn, was processed by the Precision and Recall method. According to Müller, Müller, Squire, Marchand-Maillet and Pun (2001: 1), Simulating Users is a method of obtaining relevance judgements when evaluating a CBIR system where the researcher applies human similarity judgement to simulate the intended user of the system. Müller et al. (2001: 10) also offer the method of Precision and Recall where precision equals the number of relevant documents retrieved, divided by the total number of documents retrieved and recall, which equals the number of relevant documents retrieved, divided by the total number of documents in the collection.

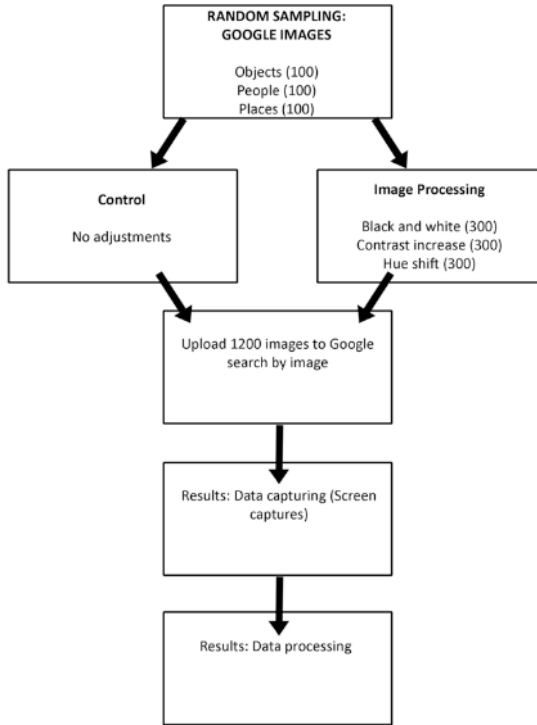
It is presented as a Precision versus Recall graph. However, it is not likely to determine the exact number of a specific image available on the Internet, nor is it according to Davis (2012: 1) feasible for a Web retrieval application like Google™ SBI to optimize for recall. This is due to the fact that most users of Google™ SBI require precision when searching by image. They are not concerned by the total number of the images available. Thus, for the purpose of this paper, only the precision value will be calculated, as this is the only factor necessary for a plagiarism detection system. One only needs to detect a single image to prove plagiarism.

## 2.2 Approach

The test images were selected through Google™ Images search engine, using simple random sampling, which, according to Vijayalakshmi and Sivapragasam (2008: 88), provides that “[e]very member of the population has an equal chance of being included in the sample”. Three separate searches, inserting the keywords “objects photography”, “people photography” and “places photography” respectively, were conducted. The three key phrases were selected, as they are standard terms used in photography to group images for a portfolio. Every tenth image was then selected until a number of 100 were reached in each group, thus resulting in a core sample of 300 images. If the tenth image was deemed inappropriate for the study, the 11th image was selected. An image was deemed inappropriate if it were not a colour image or not a photographic image.

The original images without any manipulation were tested to serve as the control group. All the images then each received a black and white conversion, contrast increase and hue shift respectively to allow for any adjustments a plagiarist may have effected in an attempt to avoid detection. The resulting 1 200 images – 300 core sample images x (3 adjustments + 1 control group) – were uploaded individually onto Google™ SBI and the results recorded via a screen capture. The recorded results were tabulated and converted to graphs for comparison to determine the feasibility of the Google™ SBI system as an effective visual plagiarism detection tool.

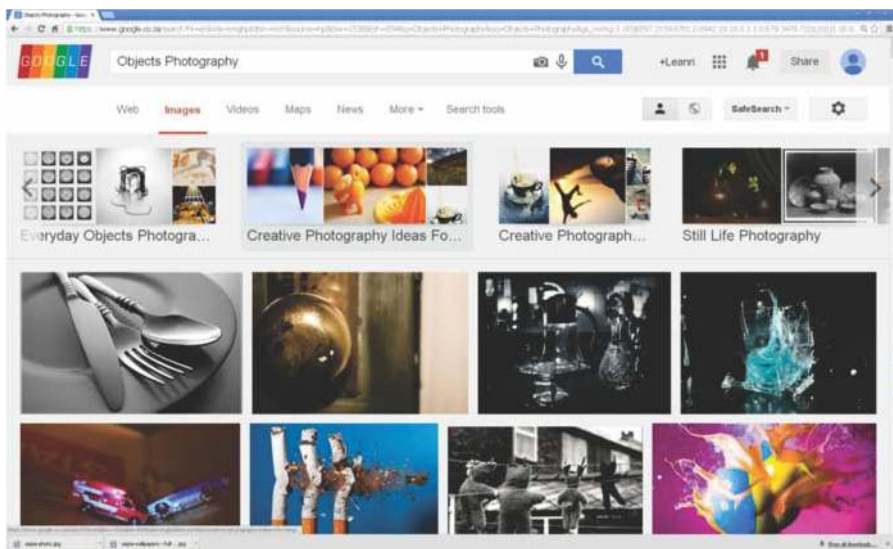
A visual depiction of the research methodology can be seen in Figure 1.



**Figure 1:** Flow chart illustrating the research methodology

### 2.3 Data Collection

The three headings used in organizing photographic portfolios, namely objects, people and places were selected to categorize the sample images. Each category heading was typed into Google™ Images search engine as the keywords “objects photography”, “people photography” and “places photography” to form the search query. One hundred images were randomly selected and saved from the search results. The first page of the results for the search using the keywords “objects photography” can be seen in Image 1.



**Image 1:** Example of results retrieved from the keywords “objects photography” as the search query (untitled screen capture)

## 2.4 Procedure

Each of the 100 images from all three categories was altered by desaturation (black and white), a 50 contrast increase and a  $-180^\circ$  hue shift respectively in Adobe® Photoshop® CS5 – the current version at the time of testing (Anon, 2010: 1) – as well as a set left unaltered to serve as the control group. Images were numbered relating to their group and the numerical order in which they were downloaded. Thus, if the image was the 67th image downloaded using the keywords “places photography”, the image was labelled as Places#67. The adjustments; desaturation, contrast increase and hue shift were performed separately on every image in the main corpus of sample images. Therefore, for each original unaltered image in the sample group there were also three copies of that image, each with a different adjustment. The images were then further labelled by adding the adjustment; in the case of Places#67, the control version of that image was labelled Places-control#67. The complete lists of sample images were as follows:

### 1. Objects

- a. 100 images unaltered – Control
- b. 100 images desaturated – Black and White
- c. 100 images with contrast increased +50 – Contrast
- d. 100 images with hue shifted  $-180^\circ$  – Hue Shift

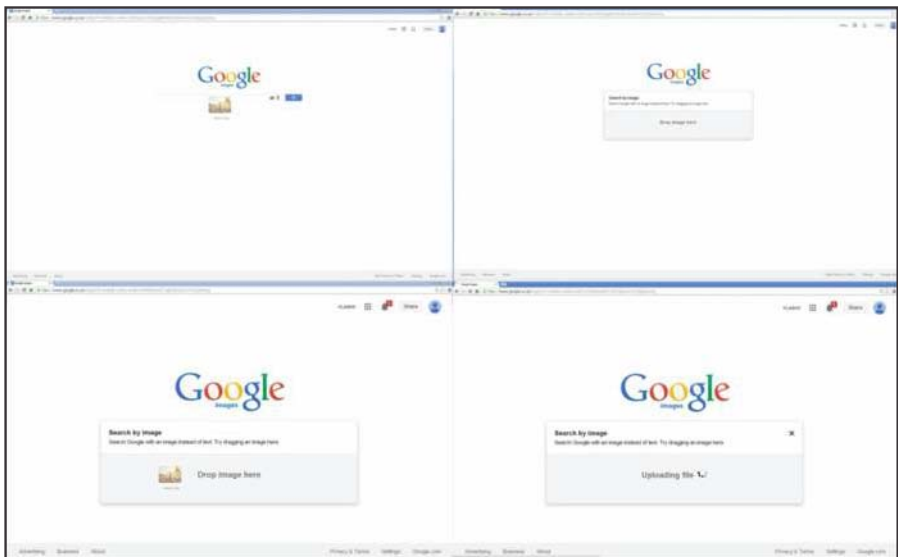
## 2. People

- a. 100 images unaltered – Control
- b. 100 images desaturated – Black and White
- c. 100 images with contrast increased +50 – Contrast
- d. 100 images with hue shifted -180° – Hue Shift

## 3. Places

- a. 100 images unaltered – Control
- b. 100 images desaturated – Black and White
- c. 100 images with contrast increased +50 – Contrast
- d. 100 images with hue shifted -180° – Hue Shift

Each of the total of 1 200 images were uploaded to Google™ SBI separately and the search results screen-captured. Image 2 shows the process of uploading Places-control#67 to Google™ Images in order to form a search query by image content.

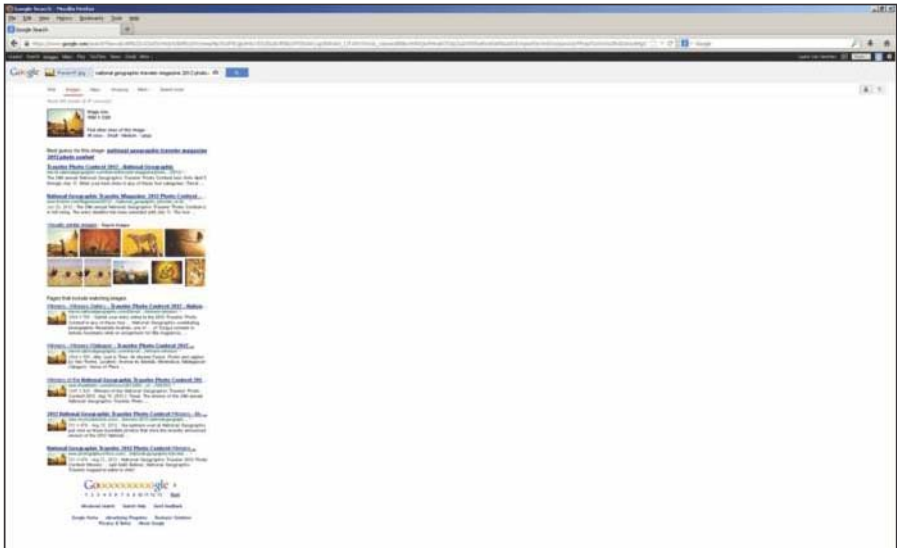


**Image 2:** Places-Control#67 being dragged and dropped into the Google™ Images search box (untitled screen capture)

The drag and drop method was used to upload the images in this study. This is achieved by clicking on the desired image and while holding in the mouse button, dragging it toward the Google™ Images search box, as seen in the top left box in Image 2.

Once Google™ Images recognises that an image is being dragged to the search box, it displays the message seen in Image 2 in the top right box: “Search by Image; search Google™ with an image instead of text. Try dragging an image here.” This message is followed by an area in which the image can be dropped, which is the result of letting go of the mouse button. In the case of dragging and dropping an image into Google™ Images, the mouse cursor will change into a thumbnail of the image selected to search with as seen in the bottom left box. The filename of the image is also attached below the image thumbnail.

After dragging the desired image file into the Google™ Images search box, the system will take a few moments to conduct its search by uploading the query image to the Google™ server and then running the image through the corpus of images on the Google™ database attempting to detect a matching image. While the search is being conducted, the search box will display the message: “Uploading file” to let the user know the search is in progress, as seen in Image 2 in the bottom right box.



**Image 3:** Results page 1 of Search by Image for Places-control#67 (untitled screen capture)

After the image has been uploaded, the results page appears, as presented in Image 3. Page 1 of the results page firstly shows the image just uploaded and indicates the other sizes in which the file is available after which the system displays its best guess for the title of the specific image. Visually similar images are then displayed as an optional link, which the user may follow.



The visually similar images are images with similar features as the uploaded image, but not a true match. The images that are true matches to the search query are listed below the visually similar images. The pages that include true matches of the image are displayed in a list that may span across multiple pages. In the case of Places-control#67, Google™ SBI retrieved 13 pages.

## 2.5 Data Analysis

The results were then studied to be tabulated. Firstly, the amount of images that were retrieved successfully was labelled as “Hits”. The total number of images retrieved were then added up and listed under the heading “Retrieved”, followed by the number of images that were matched successfully under the heading “Relevant” after which these two numbers were used to calculate the precision value as following:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

A high precision value indicates that the system returned substantially more relevant results than irrelevant results and demonstrates a CBIR system of quality.

All the results were summarized, cross-referenced and plotted as graphs into the following results:

- The image categories will be compared to determine whether the theme of the image, i.e. objects, people or places, affects the retrieval precision.
- The different adjustments made to the images, along with the control group, will be compared to establish Google™ SBI's precision in retrieving images that were converted to black and white, received a contrast increase or have undergone a hue shift.

## 3. RESULTS

During testing, the Google™ SBI system performed well and from the following results it was clear that it was a CBIR system of quality.

### 3.1 Summary of Raw Data

Following is an outline of the unprocessed results retrieved by the Google™ SBI system during testing. The objects group will be discussed, followed by the people group and then the places group. All three groups will be detailed with their adjustment results, starting with control and followed by the black and white, contrast and hue shift results respectively.

### 3.1.1 Objects Results

The results show that 92% of the images from the objects control group retrieved results. Of those 92 images that retrieved hits, only five did not achieve 100% precision. The average precision value for this group is 0.99. A number of 2 823 images were correctly matched out of a total number of 2 847 result images retrieved from the objects control group.

The objects black and white group achieved a lower average precision value of 0.94, as 11 of the 91 hits did not do so with 100% precision. Also, from the 91 of 100 images that retrieved results, 2 827 out of the total of 2 965 were correctly matched.

The objects contrast group's test results showed 92 of the 100 query images retrieved results. Eighty-eight of those images achieved 100% precision, which resulted in a 0.98 average precision score. A total number of 2 982 images were retrieved of which 2 940 were correct matches.

The objects hue shift group received 71 hits out of a possible 100. Of the total 1 731 images that were retrieved, 1 721 were matched successfully, resulting in an average precision of 0.97. Only two of the 71 hits did not score a 1.00 precision; however, both these retrievals achieved a 0.00 precision, meaning none of the images retrieved matched the query image.

### 3.1.2 People Results

The people control group's test results showed 99 of the 100 query images retrieved results. Ninety-five of those images achieved 100% precision, which resulted in a 0.99 average precision score. Of the total 5 623 images that were retrieved, 5 591 were matched successfully.

The people black and white group received 98 hits out of a possible 100. Thirty-three of the 98 hits did not score 1.00 precision with an average precision value of 0.83. A number of 5 268 images were successfully matched from a total number of 6 105 images retrieved.

The results show that 99% of images from the people contrast group retrieved results. Of the 99 images that retrieved results, only six did not achieve 100% precision. The average precision value for this group is 0.99. A total number of 5 564 images were retrieved of which 5 536 were correct matches.

The people hue shift group achieved an average precision value of 0.99, as only one of the 96 hits did not do so with 100% precision. Also, from the 96 of 100 images that retrieved results, 5 025 out of 5 053 were matched correctly.

### 3.1.3 Places Results

The results show that 98% of the images from the places control group retrieved results. Of those 98 images that retrieved hits, nine did not achieve 100% precision. The average precision value for this group is 0.97. A number of 5 732 images were matched correctly out of a total number of 5 838 result images from the places control group.

The places black and white group achieved a lower average precision value of 0.89, as 23 of the 99 hits did not do so with 100% precision. Also, from the 99 of 100 images that retrieved results, 5 016 out of the total of 5 769 were correctly matched.

The places contrast group's test results showed 98 of the 100 query images retrieved results. Eighty-nine of those images achieved 100% precision, which resulted in a 0.97 average precision score. A total number of 5 548 images were retrieved of which 5 279 were correct matches.

The places hue shift group received 86 hits out of a possible 100. Of the total of 4 421 images that were retrieved, 4 315 were matched successfully resulting in an average precision of 0.98. Six of the 86 hits did not score a 1.00 precision.

## 3.2 Processed Data

Two sets of comparisons were drawn from the results. Firstly, the three image categories (objects, people and places) were compared to see if the subject of the image could have an impact on its success rate. This was followed by a comparison of the three adjustments applied to the images (black and white, contrast and hue shift). The following variables were used to plot the graphs:  
Hits – A comparison of the total number of queries that retrieved results, whether relevant or not.

Retrieved/Relevant – A comparison of the total number of images retrieved for all the queries in that category versus the total number of images that were matched successfully to the query images.

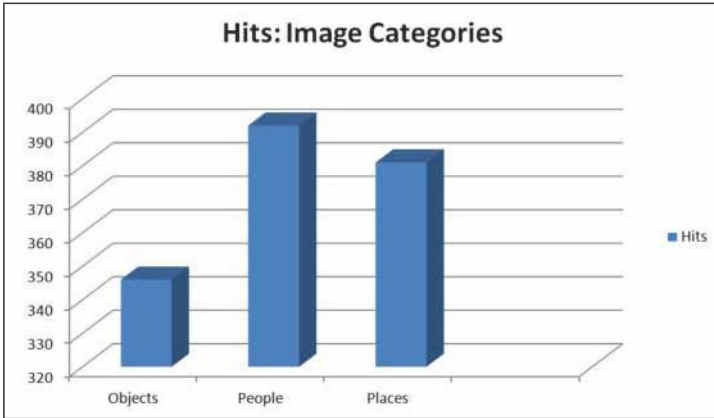
Precision – A comparison of each group's precision score average.

### 3.2.1 Image Categories Comparison

The three categories, namely people, places and objects were tabulated separately and then plotted as a graph for each of the variables.

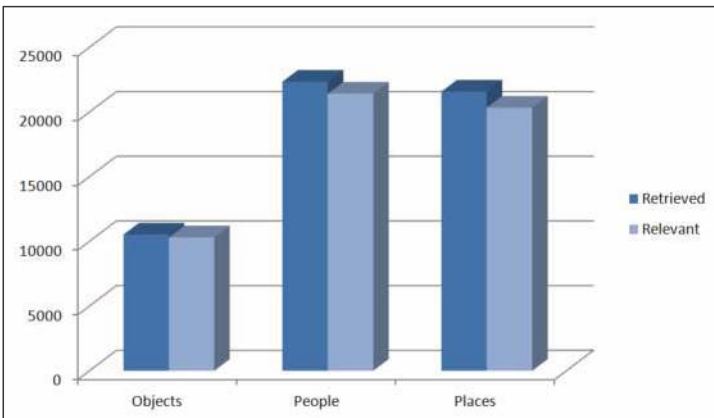
**Table 1:** Comparison of Image Categories

Hits	Retrieved	Relevant	Precision
<b>Objects: Four hundred images in total</b>			
346	10 525	10 311	0.98
<b>People: Four hundred images in total</b>			
392	22 345	21 420	0.96
<b>Places: Four hundred images in total</b>			
381	21 576	20 342	0.94



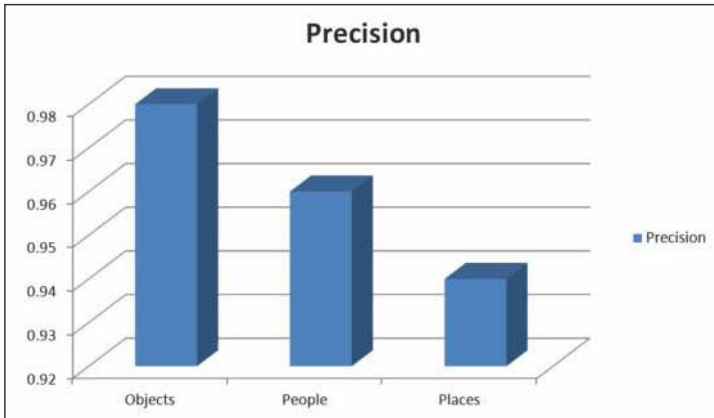
**Graph 1 – Hits: Image Categories**

Graph 1 is a visual depiction of the amount of hits from each of the image categories. It reveals that the people category successfully hit the highest amount of images, although only 11 more than places and 46 more than objects, as seen in Table 1.



**Graph 2 – Retrieved/Relevant: Image Categories**

Graph 2 shows the total amount of images retrieved against the amount of relevant images retrieved. The objects category retrieved less than half as many images as both the people and places categories, as seen in Table 1.



### Graph 3 – Precision: Image Categories

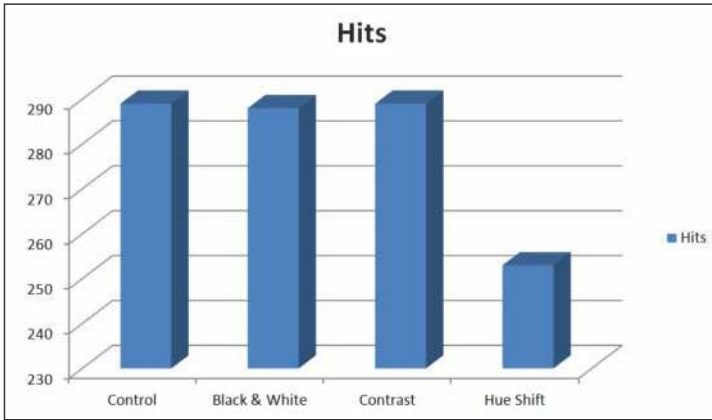
Although all three categories had a precision value in the 90th percentile, the places category scored the lowest with 0.94 and the objects the highest with 0.98, as seen in Table 1. In Graph 3 the comparison between the three categories can be seen.

#### 3.2.2 Comparison of Adjustments

The three adjustment groups, black and white, contrast and hue shift, along with the unadjusted images, the control group, were tabulated separately and then plotted as a graph for each of the variables.

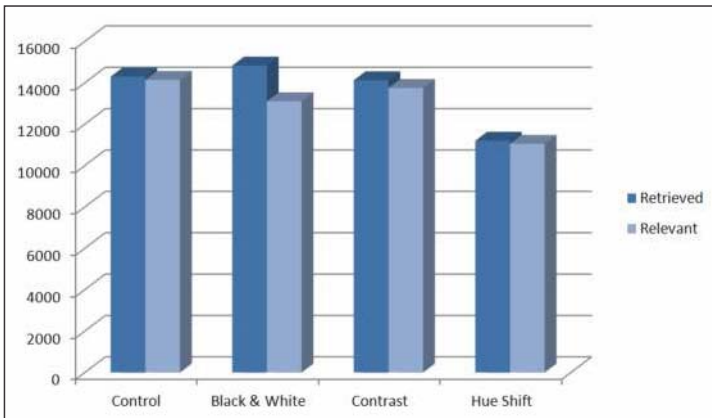
**Table 2:** Comparison of Adjustments

Hits	Retrieved	Relevant	Precision
<b>Control: Three hundred images in total</b>			
289	14 308	14 146	0.99
<b>Black and white: Three hundred images in total</b>			
288	14 839	13 111	0.88
<b>Contrast: Three hundred images in total</b>			
289	14 130	13 755	0.97
<b>Hue Shift: Three hundred images in total</b>			
253	11 205	11 061	0.99



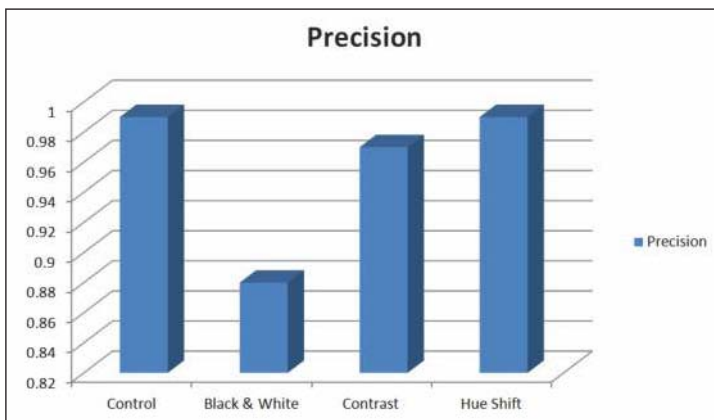
**Graph 4 – Hits: Adjustment groups**

In Graph 4 the amount of hits from the adjustment groups can be seen. The control, black and white and contrast groups received similar hits, with the hue shift group faring the worst, as seen in Table 2.



**Graph 5 – Retrieved/Relevant: Adjustment groups**

Graph 5 shows the correlation between the total amount of images retrieved and the amount of images relevant to the search query from the adjustment groups. As also seen in Table 2, the hue shift group retrieved the least of the four.



**Graph 6 – Precision: Adjustment Groups**

Graph 6 shows the precision values from the adjustment groups. The black and white group scored the lowest with a 0.88 precision value. The control and hue shift groups each scored a 0.99 precision value.

## 4. CONCLUSION

### 4.1 Overview

The problem encountered when commencing this study was that very little research had been undertaken in the detection of visual plagiarism. Also, Google™ SBI had not yet, at the commencement of this study, been systematically tested for the use of visual plagiarism detection. The objectives were to determine the percentage of sample images Google™ SBI correctly identified and the viability of the system as a visual plagiarism detection tool. The results were favourable.

### 4.2 Results

A lecturer of photography can use Google™ SBI successfully to detect images that had been appropriated unethically from the Internet and presented as a student's own work.

Google™ SBI showed an average success rate of 93.25% images retrieved, which is the result of a total of 1 119 images receiving hits out of a possible 1 200. The average precision value was 0.96, which is indicative of a high quality content-based image retrieval system. The places category scored the lowest precision value at 0.94 and the objects category the highest with 0.98.

This may be due to the fact that photographs of objects generally have simpler compositions, whereas photographs of places tend to have more complex compositions where the low-level features could easily be interchangeable in the system's view; for instance, calm water and a clear sky in a beach scene. However, even though the places category scored the lowest precision it does not indicate that the system cannot be trusted with detecting places images, as 20 342 images out of 21 576 were relevant. The places category also received 381 hits out of the total 400 images, giving it a success rate of 95.25%. The objects category only received 346 hits out of a possible 400, which gives it a success rate of 86.5%. This means that the places category had a higher image retrieval success rate, but the images retrieved in the objects category were more accurate, based on the query image. The people category scored a precision value of 0.96, placing it between the places and objects categories when it comes to accuracy. However, it had the highest success rate with 98%.

Evaluating the data from another point of view shows that the control and hue shift groups proved the most accurate with an average precision value of 0.99 each. The contrast group was almost as accurate with a 0.97 precision score, while images in the black and white group fared less well with an average precision value of 0.88. This indicated that, when retrieving images based on their content, images that had been altered by means of desaturation might retrieve less relevant images than images that had received adjustments like a contrast increase or hue shift. Images adjusted with a hue shift had an 84.3% success rate with only 253 images out of 300 receiving hits, which is the lowest success rate score for the three adjustments. The control and contrast groups received 289 out of 300 hits each, giving them the highest success rate of 96.3% with the black and white coming in at 96%, as 288 of the 300 images received hits.

### **4.3 Implications**

Based on observations made during the conducting of the research, the results of this study may have the following implications:

A lecturer of photography can use Google™ SBI successfully to detect images that had been appropriated unethically from the Internet and presented as a student's own work, even if basic adjustments were performed to alter the appearance of these images.

Google™ can develop a search by an image engine specifically designed for plagiarism detection. This system can run parallel to Google™ Scholar for the use of lecturers and other academics. This system can work with relevance feedback where the user improves the system by interacting with it.



In relevance feedback, the user uploads the query image and the system retrieves results just like in the standard process, only with relevance feedback the user is given the opportunity to mark the results as either relevant or non-relevant. The system will then use the information fed back by the user to better understand the user's needs for future queries.

Rui et al. (1998: 2) suggest also adding a keyword integration component where the user can also assign keywords to areas of an image to help the system identify tough feature areas in next searches, such as calm water and clear sky, where colour, shape and texture features are very similar. In this way, adding human interaction into the process, the system keeps getting better at retrieving images and narrowing the semantic gap. The system's automatic annotations will become so sophisticated in time that it will be able to derive high-level concepts from low-level features. This system will archive all the images that are searched, along with the images already extant in its database, similar to the archival processes of text-based plagiarism detection systems. Adding images created by students to the archives will benefit the system, as the criteria students follow (aesthetics, compositional elements and so forth) will hone the system to well-constructed images that students are more likely to plagiarise.

Google™ SBI can be used by lecturers in photography as a preventative measure. By showing students at the commencement of their studies that a system exists, which is able to detect images by their content from the Internet will discourage them from unethically appropriating images and presenting those images as their own during the course of their studies.

These recommendations may improve the photography lecturer's practice and enable him/her to spend more time engaging with the students.

## 5. BIBLIOGRAPHY

Anon. 2010. Adobe Photoshop Release History [Internet]. Available from: [http://kb2.adobe.com/community/publishing/925/cpsid\\_92587/attachments/photoshop\\_release\\_versions\\_history.pdf](http://kb2.adobe.com/community/publishing/925/cpsid_92587/attachments/photoshop_release_versions_history.pdf) [Accessed 30 November 2015].

Datta, R., Joshi, D., Li, J. & Wang, J.Z. 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 2(40): 1-60, April.

Davis, J. 2012. Recall & Precision [Internet]. Available from: <http://www.drjasondavis.com/2012/01/21/recall-precision/> [Accessed 30 November 2015].

Garrett, L. & Robinson, A. 2012. Spot the Difference! Plagiarism Identification in the Visual Arts [Internet]. Available from: [http://www.research.ucreative.ac.uk/1194/1/ewic\\_ev12\\_s2paper1.pdf](http://www.research.ucreative.ac.uk/1194/1/ewic_ev12_s2paper1.pdf) [Accessed 30 November 2015].

Gutenko, G. 2000. OPEN VS CLOSED: Academic Dishonesty Issues in Digital Production Course Instruction [Internet]. Available from: <http://files.eric.ed.gov/fulltext/ED454558.pdf> [Accessed 30 November 2015].  
Howard, R.M. 2007. Understanding Internet Plagiarism. Elsevier, 24(2007): 3-15, 2007.

Müller, H., Müller, W., Squire, D.M., Marchand-Maillet, S. & Pun, T. 2001. Performance Evaluation in Content-based Image Retrieval: Overview and Proposals. Science Direct, (22)5: 593-601, April.

Rui, Y., Huang, T.S., Ortega, M. & Mehrotra, S. 1998. Relevance Feedback: A Power Tool for Interactive Content-based Image Retrieval [Internet]. Available from: <http://research.microsoft.com/en-us/um/people/yongrui/ps/csvt98.pdf> [Accessed 30 November 2015].

Stoltenkamp, J. and Kabaka, M. (2014) Turnitin Adoption and Application at a HEI: A Developmental Approach. Creative Education, 5, 1043-1052.

Vadivel, A., Majumdar, A.K. & Sural, S. 2004. Characteristics of Weighted Feature Vector in Content-based Image Retrieval Applications. Intelligent Sensing and Information Processing, 1(1): 127-132.

Vijayalakshmi, G. & Sivapragasam, C. 2008. Research Methods: Tips and Techniques. Chennai: MJP Publishers.