

# The Understanding of Exchangeable Image File (Exif) Metadata of Images: Towards Disseminating the Awareness to the Society

Shafinaz Mohammad Niyaz Khan<sup>a</sup>, Juzlinda Mohd Ghazali<sup>b</sup>, Lailatul Qadri Zakaria<sup>c</sup>, Norzalina Zainudin<sup>d</sup>, Siti Noor Ahmad<sup>e</sup> & Khairil Ashraf Elias<sup>f</sup>

<sup>a, b, d, e, f</sup>*International Islamic University College Selangor, Selangor, Malaysia*  
<sup>c</sup>*Universiti Kebangsaan Malaysia, Selangor, Malaysia*

**Abstract.** Most people nowadays, prefer to upload or store their personal digital images on social media or websites, but how far do they know about what kind of metadata have been stored behind those digital images that they are revealing to other people who can access the Internet? How aware are the lecturers and students of Faculty of Science and Information Technology (FSTM) of this issue in order to save the society from being misled through posting images by using the social media and websites? This research is meant to study on how far does the knowledge of EXIF metadata of digital images is being acquired and understood among the lecturers and students of FSTM in order for them to educate the society. This is to make aware of society's behaviors, rights and protections on personal information through their digital images in the Internet world. The findings of this research show that there is a clear positive correlation between having the knowledge of EXIF metadata among the respondents with disseminating the awareness of it to the society. On the other hand, there is a negative correlation between having the knowledge of digital images among the respondents with disseminating the awareness of EXIF metadata of digital images to the society. It could be concluded that those with the knowledge of EXIF metadata of digital images could possibly disseminate the awareness of it to the society as for awareness and concern purposes, whereas those with the knowledge of digital images but without the knowledge of EXIF metadata of digital images could possibly be unable to disseminate the awareness of EXIF metadata of digital images to the society.

**Keywords:** digital images, EXIF metadata.

## 1.0 INTRODUCTION

EXIF metadata is data about data that is automatically embedded into a digital image. Running a quantitative survey on the lecturers and students of FSTM, International Islamic University College Selangor (KUIS) in their acquired knowledge of EXIF metadata of digital images could establish a benchmark in knowing whether or not they could disseminate the awareness of EXIF metadata of digital images to the society as for awareness and concern purposes towards the society.

The idea of this research arises due to the concern of the researcher towards the increase of cybercrimes, which are computer related crimes in our local nation as has been reported by the nation's Cybersecurity Malaysia and the News Straits Times. According to Cybersecurity Malaysia, the statistics of cybercrimes from the year of 2012 is 9986 cases, 2013 with 10636 cases, 2014 with 11918 cases, 2015 with 9915 cases and last year with 8334 cases. Even though there is a decrease rate of cybercrimes from 2014 to 2016, the total amount of 8334 cases in 2016 is still a big figure and must be reduced.

Geo location data, which is a one of the EXIF metadata category, are embedded onto digital images when the user does not disable it. An article from Exposing The Invisible regarding "Digging Into Digital Images: Extracting Batch Location Data Automatically" reveals that the data such as Global Positioning System (GPS) Altitude, GPS Date/Time, GPS Latitude, GPS Longitude, and GPS Position are among the geo location data that are embedded onto the digital images. This could be dangerous towards some people, groups or families.

In an article of Wired Magazine by Mathew Honan (2009), he stated the following statement "I ran a little experiment. On a sunny Saturday, I spotted a woman in Golden Gate Park taking a photo with a Third Generation (3G) iPhone. Because iPhones embed geodata into photos that users upload to Flickr or Picasa, iPhone shots can be automatically placed on a map. At home I searched the Flickr map, and score – a shot from today. I clicked through to the user's photostream and determined it was the woman I had seen earlier. After adjusting the settings so that only her shots appeared on the map, I saw a cluster of images in one location. Clicking on them revealed photos of an apartment interior – a bedroom, a kitchen, a filthy living room. Now I know where she lives." Isn't this scary enough?

According to the local News Straits Times, on 21<sup>st</sup> March 2017, around 2:10pm at Kuala Terengganu, Terengganu, Malaysia, a 2015 survey of school children in Malaysia, conducted by the police's sexual, women and child investigation unit, and telecommunications company Digi's CyberSAFETM, has revealed shocking statistics. Their findings show that one out of ten children is likely to have been asked to upload intimate photos or videos of themselves onto the internet. There has been a 300% increase in internet-initiated rape cases in the country between 2010 and 2015, and that is close to 80% of reported rape cases by predators in Malaysia over the past two years involved internet acquaintances, with the majority of victims below 18 years of age.

This research could also show how important the relation between having the knowledge and disseminating the awareness to the society as this could lead the society to a better Information and Communication Technology (ICT) knowledgeable informed society. The potential of using this research to create a benchmark of FSTM lectures and students' progress quality output in order to make and improve future progress of teaching and learning skills, as well as in the society is yet been explored and deemed promising.

## 2.0 LITERATURE REVIEW

Alonso (n.d.), Sun (2002), Snavely (2007), Smith (2012), and Sumner (2014) mentioned that EXIF metadata are embedded on all digital images automatically by digital devices that captured the images. Alonso (n.d.) describes that one could distinguish whether a digital image has been altered or edited from its original form by checking through the EXIF metadata of the image. EXIF metadata can provide useful benefits for its user such as the image's authors name identification, the computer that was been used to edit or store the image, the type of software that was been used to manipulate the image, the date of when the image was been taken or edited and for many other beneficial properties (Alonso, n.d. and Sun et al., 2002).

EXIF metadata could reveal useful information about the image (Sun et al., 2002; Snavely et al., 2007; and Sumner, 2014). It can reveal the information settings of a digital camera before the image was been taken. By this way, a photographer could learn or practice taking digital photos from a set of camera settings that the photographer would like to repeat based on the given EXIF metadata of the digital image which the photographer adores. There are social media networks, cloud storages, and websites that keeps EXIF metadata of digital images in their database system but not revealing it to the public (Smith et al., 2012) but could be accessed by hackers.

EXIF tag which is a type of metadata that is being tagged to an image by the owner or user of the digital image could bring ease and fast recovery of the type of image that a person is searching for (Snavely et al., 2007). It could also bring harm to its owner or user from a hacker who may download images with specific tags and edit them to give a misinterpreted image with wrongly thoughts or misled the edited image with big negative issue that could create chaotic scenarios towards the society. For example, a digital image that was uploaded to a social media about a neighbour having a nice and good welfare party in his house, and he tagged the picture with a tag name "welfare party" but was been downloaded by a hacker. The hacker then edit the picture to make it look as if the welfare party was conducting drug abuses in it and uploaded the edited image back again onto the owner's social media. Afterwards, his friends and neighbourhood will get to know of the new edited pictures and have negative thoughts towards him and this could create an ugly social impact towards him.

Geo location data that is either automatically or manually implemented onto digital image is also a type of EXIF metadata. There are many kinds of geo metadata under the geo location EXIF metadata category such as GPS Latitude Ref, GPS Longitude Ref, GPS Altitude Ref, GPS Time Stamp, GPS Img Direction Ref, GPS Img Direction, GPS Date Stamp, GPS Altitude, GPS Date/Time, GPS Latitude, GPS Longitude, and GPS Position (Exposing The Invisible, n.d.). According to an article by the United States Army, the fact that most of the digital cameras nowadays does not implement geo location automatically is not always true. However, the owner or user of the camera should study and know the functions of the camera that they will use and make sure that the GPS function is being turn off. Even though if there is an image with a geo location metadata embedded in it, the user should learn and know on how to use

certain software to remove those metadata before uploading them onto the website or any social media for security purposes.

Thus, educators and youths plays an important role in the society in knowing and preventing misled usage of EXIF metadata in digital images. They will be the ones who could shape the society to become a better ethical society that will form a better respected nation. A research survey on the knowledge awareness of lecturers and students of FSTM, KUIS on EXIF metadata in digital images is just the beginning or the stepping stone in shaping the society and nation as a whole. The following topic is about Research Methodology, which will describe the theoretical analysis of the methods that are been applied to this research in order to gather the required data to be manipulated to form meaningful information.

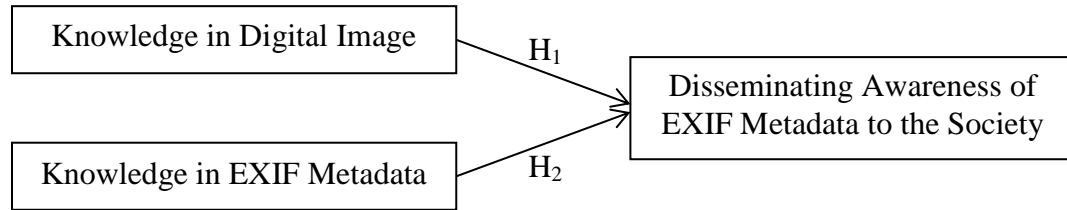
### **3.0 RESEARCH METHODOLOGY**

A survey on targeted respondents of the research through printed forms of questionnaire development, distribution and collection of answered questionnaires (data collection) is the methodology implementation for this research. The targeted subjects will be of any 15 participants of FSTM's lecturers from a total of 28 lecturers, and any of 15 participants of FSTM's students, consist of male and female genders, mentioning on how long they have worked or studied under FSTM and from which department or program that they belong to. 15 participants from each group of FSTM's lecturers and students (a total of 30 participants) would be enough as for the research is meant to tabulate a benchmark from a small group of participants and FSTM consists of not many lecturers and students.

The questions of the questionnaire is in English and Bahasa Melayu, in a form of selections by placing a tick at the appropriate and relevant place, stating some descriptions, and choosing a scale out of five scales from "Absolutely do not know" which represents a very weak feedback towards the question, to "Absolutely know" which shows a very strong belief or feedback on the question. The questions in the questionnaire form are in the scale of nominal (example, type of respondent, which is either lecturer or student, and gender, which is male or female), ordinal (example, student's semester), interval (example, scale of knowing for each questions in Section C), and ratio (example, designation of lecturer, and course of student). They are the very basic questions in gathering data for this research. The type of questions in the questionnaires will be on what, where, why, who and how EXIF metadata of digital images relates with its purposes, functions, advantages, disadvantages and methods on how to overcome such situation towards its subjects or participants.

The researcher applies simple random sampling method for the sampling target of FSTM lecturers and students. The data collections from the questionnaire form in Section C of part I represents the first independent variable which is the knowledge in digital image, and Section C of part II represents the second independent variable which is the knowledge in EXIF metadata. Section C of part III represents the dependent variable which is disseminating awareness of EXIF metadata of digital images to the society. This

could signify the testing of the objective or purpose of this research which is to see the relationship between the two independent variables which are knowledge in digital image and knowledge in EXIF metadata with the dependent variable which is disseminating awareness of EXIF metadata of digital images to the society in order to create awareness of it. The theoretical framework of this research is shown in the figure (Figure 1) below:



**FIGURE 1:** The theoretical framework of this research.

The researcher uses IBM SPSS Statistics software to digest the gathered data from the answered questionnaire forms, and tabulate different kinds of informative data, which can provide meaningful information for this research in order to achieve the objective or purpose of this research.

#### 4.0 RESULTS AND DISCUSSIONS

In order to view the outcome from the answered questionnaires, the researcher had used IBM SPSS Statistics software to tabulate the gathered data in a meaningful manner and produce output to present useful information for this research. All feedbacks from the respondents are being entered into the software and various output instructions were being entered in order for the software to provide the calculations, statistics, graphs, charts, and tables for the researcher to generate useful information.

From the distributed questionnaires, the researcher had gathered 30 feedbacks from two groups of respondents whereby 15 answered questionnaire forms are from FSTM's lecturers and another 15 are from FSTM's students. All questionnaire forms were distributed accordingly by using and implementing the methodology of this research's sampling method, which had been explained in the previous subtopic. The number of questionnaires with non-missing value is shown in Table 1 below:

**TABLE 1:** The number of distributed and gathered questionnaire forms.

|   |         | Statistics       |             |
|---|---------|------------------|-------------|
|   |         | Lecturer/Student | Male/Female |
| N | Valid   | 30               | 30          |
|   | Missing | 0                | 0           |

Table 2 shows the amount and percentage of lecturers and students who answered the questionnaire form. From here, 15 lecturers out of 15 lecturers and 15 students out of

15 students responded to the questionnaire forms, giving a 100% of answered questionnaire forms from which are been distributed.

**TABLE 2:** The number and percentages of lecturers and students who answered the questionnaire form.

| Lecturer/Student |          | Frequency | Percent | Valid Percent | Cumulative Percent |
|------------------|----------|-----------|---------|---------------|--------------------|
| Valid            | Lecturer | 15        | 50.0    | 50.0          | 50.0               |
|                  | Student  | 15        | 50.0    | 50.0          | 100.0              |
|                  | Total    | 30        | 100.0   | 100.0         |                    |

Table 3 shows the amount and percentage of individual gender type who answered the questionnaire form. From the respondents of lecturers and students, 10 questionnaire forms were been answered by the male gender and 20 by the female gender, giving a 33.3% for male and 66.7% for female out of 100%.

**TABLE 3:** The number and percentages of individual gender type who answered the questionnaire form.

| Male/Female |        | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------------|--------|-----------|---------|---------------|--------------------|
| Valid       | Male   | 10        | 33.3    | 33.3          | 33.3               |
|             | Female | 20        | 66.7    | 66.7          | 100.0              |
|             | Total  | 30        | 100.0   | 100.0         |                    |

We had also run the following statistics (as shown in Table 4 below) from the SPSS program in order to get the reading of the number of years that the lecturers who had responded to the questionnaire form had worked under FSTM, whereby the minimum number of years are from one to two years and the maximum number of years is 19 years and above. On the other hand, the table shows the number of semesters that the students who had responded to the questionnaire form had been studying under FSTM, with the minimum semester of seven and the maximum semester of eight. The table also tabulates the minimum and maximum scale of answer for all questions in Section C by both lecturers and students, and also providing the average scale for each answer which can be seen under the “Mean” column. The scale of 1.00 represents “Absolutely do not know”, 2.00 represents “Do not know”, 3.00 represents “Not sure”, 4.00 represents “Know”, and 5.00 represents “Absolutely know”.

**TABLE 4:** Number of years, number of semesters, minimum and maximum scale of answer as well as its average answer scale for all questions in Section C by both lecturers and students.

**Descriptive Statistics**

|                           | N  | Minimum | Maximum | Mean   | Std. Deviation |
|---------------------------|----|---------|---------|--------|----------------|
| Years of Being A Lecturer | 15 | 1.00    | 10.00   | 6.6667 | 2.55417        |
| Student's Semester        | 15 | 7.00    | 8.00    | 7.6667 | .48795         |
| Section C, I, Q1          | 30 | 1.00    | 5.00    | 4.2000 | .76112         |
| Section C, I, Q2          | 30 | 1.00    | 5.00    | 4.0333 | .80872         |
| Section C, I, Q3          | 30 | 2.00    | 5.00    | 4.0667 | .82768         |
| Section C, I, Q4          | 30 | 1.00    | 5.00    | 4.0333 | .85029         |
| Section C, I, Q5          | 30 | 2.00    | 5.00    | 4.0000 | .78784         |
| Section C, II, Q1         | 30 | 1.00    | 4.00    | 2.7667 | 1.00630        |
| Section C, II, Q2         | 30 | 1.00    | 5.00    | 2.7667 | .97143         |
| Section C, II, Q3         | 30 | 1.00    | 4.00    | 2.4000 | .81368         |
| Section C, II, Q4         | 30 | 1.00    | 4.00    | 2.5000 | .93772         |
| Section C, II, Q5         | 30 | 1.00    | 4.00    | 2.3667 | .80872         |
| Section C, III, Q1        | 30 | 1.00    | 5.00    | 2.8667 | 1.07425        |
| Section C, III, Q2        | 30 | 1.00    | 5.00    | 2.9667 | .99943         |
| Section C, III, Q3        | 30 | 1.00    | 4.00    | 2.7667 | .97143         |
| Section C, III, Q4        | 30 | 1.00    | 4.00    | 2.6667 | .84418         |
| Section C, III, Q5        | 30 | 1.00    | 5.00    | 3.1333 | 1.04166        |
| Valid N (listwise)        | 0  |         |         |        |                |

From the regression analysis that had been tabulated, which forms the coefficients table below (Table 5), the estimated regression equation can be explained as follow:

$$\text{Disseminating Awareness of EXIF Metadata to the Society} = 0.993 - 0.011 \text{ Knowledge in Digital Image} + 0.755 \text{ Knowledge in EXIF Metadata.}$$

Given from the table below (Table 5), DisseminatingAwarenessOfEXIFMetadataToTheSociety represents the Dependent Variable of Disseminating Awareness of EXIF Metadata to the Society, KnowledgeInDigitalImage represents Independent Variable I of Knowledge in Digital Image, KnowledgeInEXIFMetadata represents Independent Variable II of Knowledge in EXIF Metadata. The above equation describes in a very specific term on how the two independent variables (KnowledgeInDigitalImage and KnowledgeInEXIFMetadata) are related to disseminating awareness of EXIF metadata to the society.

The researcher also concludes that the explanations on the relationship between each independent variable and dependent variable are as follow:

- From the b value for KnowledgeInDigitalImage of -0.011, knowledge in digital image has an inverse relationship with disseminating awareness of EXIF metadata to the society. An increase in knowledge in digital image is expected to decrease

the disseminating awareness of EXIF metadata to the society. In other words, a one unit increase in knowledge in digital image will decrease the disseminating awareness of EXIF metadata to the society by 0.011 units.

- ii. From the b value for KnowledgeInEXIFMetadata of 0.755, knowledge in EXIF metadata has a positive relationship with disseminating awareness of EXIF metadata to the society. An increase in knowledge in EXIF metadata is expected to increase the disseminating awareness of EXIF metadata to the society. In other words, a one unit increase in knowledge in EXIF metadata will increase the disseminating awareness of EXIF metadata to the society by 0.755 units.

**TABLE 5:** Regression analysis and the coefficients table.

**Variables Entered/Removed<sup>a</sup>**

| Model | Variables Entered                                               | Variables Removed | Method |
|-------|-----------------------------------------------------------------|-------------------|--------|
| 1     | KnowledgeInEXIFMetadata<br>KnowledgeInDigitalImage <sup>b</sup> |                   | Enter  |

a. Dependent Variable:  
DisseminatingAwarenessOfEXIFMetadataToTheSociety

b. All requested variables entered.

**Coefficients<sup>a</sup>**

| Model |                         | Unstandardized Coefficients |            | Standardized Coefficients | t     | Sig. |
|-------|-------------------------|-----------------------------|------------|---------------------------|-------|------|
|       |                         | B                           | Std. Error | Beta                      |       |      |
| 1     | (Constant)              | .993                        | .614       |                           | 1.619 | .117 |
|       | KnowledgeInDigitalImage | -.011                       | .139       | -.010                     | -.082 | .935 |
|       | KnowledgeInEXIFMetadata | .755                        | .122       | .771                      | 6.204 | .000 |

a. Dependent Variable: DisseminatingAwarenessOfEXIFMetadataToTheSociety

**Residuals Statistics<sup>a</sup>**

|                      | Minimum  | Maximum | Mean   | Std. Deviation | N  |
|----------------------|----------|---------|--------|----------------|----|
| Predicted Value      | 1.6915   | 3.9680  | 2.8800 | .62140         | 30 |
| Residual             | -1.20161 | .84738  | .00000 | .51628         | 30 |
| Std. Predicted Value | -1.913   | 1.751   | .000   | 1.000          | 30 |
| Std. Residual        | -2.246   | 1.584   | .000   | .965           | 30 |

a. Dependent Variable: DisseminatingAwarenessOfEXIFMetadataToTheSociety

From the regression analysis, which also forms the ANOVA table below (Table 6), the explanations on the significance of the overall regression model are as follow:

H<sub>0</sub>: The overall regression model will not significantly fit the data.

H<sub>1</sub>: The overall regression model will significantly fit the data.

F = 19.557; df = 2,27; p = 0.0001



The researcher rejects  $H_0$  if  $p < 0.05$ . Therefore, the researcher rejects  $H_0$ . The overall model with the two predictors of knowledge in digital image, and knowledge in EXIF metadata has significantly explained the variation in disseminating awareness of EXIF metadata to the society.

**TABLE 6:** The ANOVA table.

**ANOVA<sup>a</sup>**

| Model |            | Sum of Squares | df | Mean Square | F      | Sig.              |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1     | Regression | 11.198         | 2  | 5.599       | 19.557 | .000 <sup>b</sup> |
|       | Residual   | 7.730          | 27 | .286        |        |                   |
|       | Total      | 18.928         | 29 |             |        |                   |

a. Dependent Variable: DisseminatingAwarenessOfEXIFMetadataToTheSociety

b. Predictors: (Constant), KnowledgeInEXIFMetadata, KnowledgeInDigitalImage

From the regression analysis that also forms the Model Summary table, which is shown in the table (Table 7) below, shows that the R square for the overall regression model is 0.592. In other words, 59.2% of the variation in disseminating awareness of EXIF metadata to the society is explained by knowledge in digital image, and knowledge in EXIF metadata.

**TABLE 7:** The Model Summary table.

**Model Summary<sup>b</sup>**

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | .769 <sup>a</sup> | .592     | .561              | .53506                     |

a. Predictors: (Constant), KnowledgeInEXIFMetadata, KnowledgeInDigitalImage

b. Dependent Variable: DisseminatingAwarenessOfEXIFMetadataToTheSociety

From the research output information, the researcher has been able to achieve the objective of the research which is to acquire feedbacks from the lecturers and students of FSTM, KUIS on the knowledge of digital images and EXIF metadata that exists behind digital image in disseminating the awareness on the existence of EXIF metadata of digital images to the society in order to know the relationship between knowing the knowledge of digital images and knowledge of EXIF metadata of digital images with disseminating awareness of it to the society.

The objective had been achieved by running the questionnaire distributions among the respondents who are lecturers and students of FSTM and gathering fully answered feedbacks from the respondents through those questionnaire forms. From here, the researcher keyed-in the gathered data from the respondents into the IBM SPSS Statistics software and producing a tabulation output out of it, which gives rational information pertaining what the researcher would like to know from this research. However, the list of questions under Section C I, which are I.1, I.2, I.3, I.4, and I.5 may

not be suitable for this research as to find the relations with disseminating the awareness of EXIF metadata of digital images to the society may not be shown by respondents who has knowledge of digital image.

Even through digital image has something to do with EXIF metadata, questions pertaining digital image on its own may not be suitable but questions pertaining digital image with EXIF metadata may be suitable for this research, in exchange for question I.1, I.2, I.3, I.4, and I.5.

## 5.0 CONCLUSION

The researcher can conclude that lecturers and students of FSTM, KUIS with a given knowledge of digital image will not necessary be able to disseminate the awareness of EXIF metadata of digital images to the society but lecturers and students of FSTM, KUIS with a given knowledge in EXIF metadata might be able to disseminate the awareness of EXIF metadata of digital images to the society, and this is what can be obtained from the research that has been conducted.

## REFERENCES

- Alonso, C., E. Rando, F. Oca and A. Guzmán. N.d.. “*Disclosing Private Information from Metadata, Hidden Info and Lost Data*,” Black Hat, London, United Kingdom. Pg 6, 16-17.
- David, A. 2017. “*Survey Reveals Alarming Statistics on Internet-Initiated Sex Abuse of M'sian Minors*,” New Straits Times. Retrieved April 7, 2017, from <http://www.nst.com.my/news/2017/03/223005/survey-reveals-alarming-statistics-internet-initiated-sex-abuse-msian-minors>
- Exposing The Invisible, n.d.. “*Digging Into Digital Images: Extracting Batch Location Data Automatically*,” Retrieved May 2, 2017, from <https://exposingtheinvisible.org/resources/obtaining-evidence/image-digging>
- Honan, M.. 2009. “*I Am Here: One Man's Experiment With the Location-Aware Lifestyle*,” Wired. Retrieved May 2, 2017, from I Am Here: One Man's Experiment With the Location-Aware Lifestyle <https://www.wired.com/2009/01/lp-guineapig/>
- Malaysia Computer Emergency Response Team, 2017. “*MyCERT Incident Statistics*,” Cybersecurity Malaysia, Selangor Darul Ehsan, Malaysia.
- Smith, M., C. Szongott, B. Henne and G. V. Voigt. 2012. “*Big Data Privacy Issues in Public Social Media*,” Leibniz Universitat Hannover and L3S Research Center, Hannover, Germany. Pg 3.
- Snavely, N., S. M. Seitz and R. Szeliski. 2007. “*Modeling the World from Internet Photo Collections*,” Springer Science+Business Media, LLC 2007, New York, United States of America. Pg 3, 5, 7.

Sumner, R.. 2014. “*Processing RAW Images in MATLAB*,” Department of Electrical Engineering, University of California, Santa Cruz, United States of America. Pg 1.

Sun, Y., H. Zhang, L. Zhang and M. Li. 2002. “*MyPhotos – A System for Home Photo Management and Processing*,” Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China. Pg 1.

Zainudin, N, n.d.. “*Research Methodology in Business*,” Module, International Islamic University College Selangor (KUIS), Selangor, Malaysia. Pg 33-188.