ENGINEERING JOURNAL

Article

Location Estimation of a Photo: A Geo-signature MapReduce Workflow

Pakpoom Mookdarsanit[a] **and Montean Rattanasiriwongwut**[b,*]

Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand
E-mail: [a]pakpoom.mookdarsanit@gmail.com, [b]montean.r@it.kmutnb.ac.th (Corresponding author)

Abstract. Location estimation of a photo is the method to find the location where the photo was taken that is a new branch of image retrieval. Since a large number of photos are shared on the social multimedia. Some photos are without geo-tagging which can be estimated their location with the help of million geo-tagged photos from the social multimedia. Recent researches about the location estimation of a photo are available. However, most of them are neglectful to define the uniqueness of one place that is able to be totally distinguished from other places. In this paper, we design a workflow named G-sigMR (Geo-signature MapReduce) for the improvement of recognition performance. Our workflow generates the uniqueness of a location named Geo-signature which is summarized from the visual synonyms with the MapReduce structure for indexing to the large-scale dataset. In light of the validity for image retrieval, our G-sigMR was quantitatively evaluated using the standard benchmark specific for location estimation; to compare with other well-known approaches (IM2GPS, SC, CS, MSER, VSA and VCG) in term of average recognition rate. From the results, G-sigMR outperformed previous approaches.

Keywords: Photo location estimation, location uniqueness, location recognition, MapReduce structure.

ENGINEERING JOURNAL Volume 21 Issue 3
Received 3 October 2016
Accepted 21 December 2016
Published 15 June 2017
Online at http://www.engj.org/
DOI:10.4186/ej.2017.21.3.295

DOI:10.4186/ej.2017.21.3.295

1. Introduction

Nowadays, influence of social multimedia has rapidly changed the world. Social multimedia is a virtual world that enables users to share their articles, photos, videos, etc. Social users have built a large number of information stored at the datacenter. IDC Digital Universe forecasted the amount of information over the social multimedia will be 40ZB within 2020 [1]. Since there are a large number of information, especially in type of photography that are easily taken from the mobile devices; and quickly shared over the social multimedia during their travels. Million photos on the social multimedia are tagged with their geo-locations (also called geo-tagged photos) which are useful for image retrieval [2] in terms of browsing, searching, mining, and organizing. However, there are many photos without geo-tagging (also called geo-untagged photos) that are shared on the social multimedia. The EXIF data is automatically tagged during the capturing a photo for description of the location. However, the photos with enhancement process (such as cropping, blending, sharpening and other effects) are easy for loss of EXIF data. To that end, it is feasible that some geo-untagged photos can be estimated their locations with the help of another million geo-tagged photos from the large-scale dataset. The methodology for finding the geographical location of a photo (where it was taken) is called "location estimation of a photo" which is a new branch of image retrieval. Visual content (such as color, texture, and shapes) and other textual metadata (such as annotations, tags, duplicated comments and/or previous user's sharing) of a geo-untagged photo can be used to retrieve any similar scenes from the dataset. The location of geo-untagged photo is estimated by the location of the most similar geo-tagged photo (or the most similar group that is categorized from geo-tagged photos) from the dataset. In 2008, the first groundwork was found by Hays and Efros [3]. They developed a simple image retrieval for automatic geo-tagging of their photos (well-known as IM2GPS) that they were taken with the quote-worthy question as "What can you say about where these photos were taken?"

Later, most researches about the location estimation of a photo still focused on the recognition performance by a traditional spatial coding (named SC) [4] and an indexing of hierarchical global feature clustering with local feature refinement under the measurement of cosine-based similarity (named CS) [5]. Some approaches adapt the visual synonym as in the maximally stable extremal region (MSER) algorithm for the salient region mining [6]. Recently, the visual spatial contents arrangement (named VSA) [7] and the group of visual contents (named VCG) [8] from a photo were built to matched the similar scenes from the dataset in 2014 and 2015, respectively.

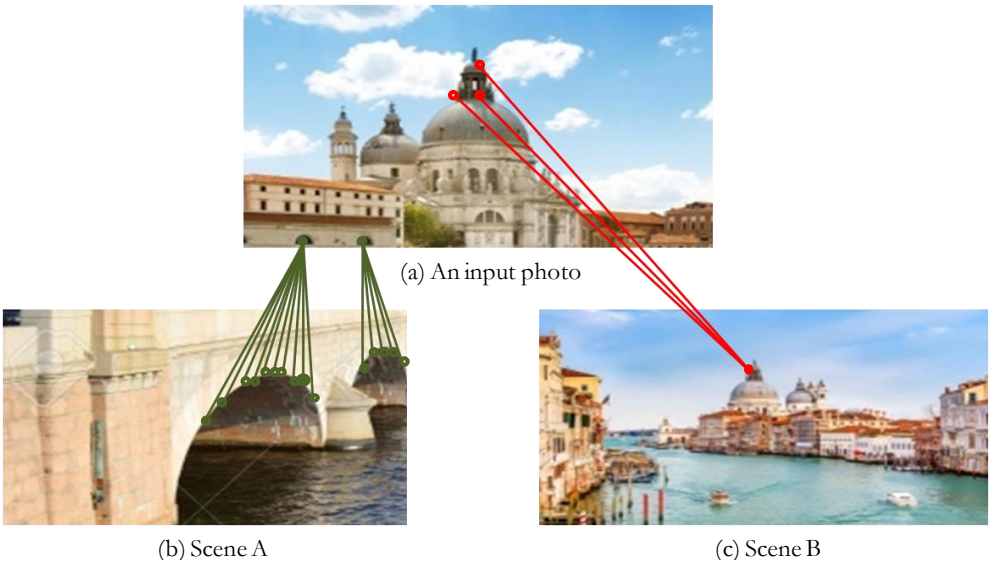

Fig. 1. In case of the error from similarity-based location estimation of a photo.

However, those previous approaches have a common ignorance. Since their location estimations are computed from a geo-tagged photo or a group of geo-tagged photos from the dataset that have the highest number of similar visual contents with the geo-untagged photo. This idea is found that it is easy to produce the error because the most similarity sometimes cannot directly estimate the target location of a photo.

As in Fig. 1, the geo-untagged photo is used to find the similar scenes from the dataset which are retrieved as the result of scene A (St. Petersburg, Russia) and B (Venice, Italy). Since the geo-untagged photo was taken at Venice as the same location as the scene B. Unfortunately, the scene A has a more number of similar contents than the scene B. Hence, location of the geo-untagged photo is wrongly estimated as taken from St. Petersburg as the same location as scene A.

Location estimation of a photo needs a more procedure to define the uniqueness of a place that to be distinguished from other places. The contribution of our paper is to design a workflow named G-sigMR (Geo-signature MapReduce) for the improvement of recognition performance. First, the important visual contents are selected from a geo-untagged photo (in term of a vector). Second, the system applies the Map-Reduce indexing to map only the similar scenes from the millions of geo-tagged photos in the large-scale dataset; and store them in the candidate space. In contrast, some irrelevant scenes are reduced. Third, all scenes in the candidate space are grouped together. The scenes taken from the same location are grouped into the same group. Fourth, each group is defined the uniqueness of the group named Geo-signature which is summarized from their visual synonyms. Finally, location of the geo-untagged photo is estimated by the group that has the maximal signature. Our G-sigMR is compared to the previous approaches [3-8] based on Geo-tagged large-scale dataset (GOLD) [9-10] which contains 3.3 millions of geo-tagged photos. From the results, our G-sigMR had the average recognition rate more than 90% which outperformed other approaches.

The remainder of this paper is organized as follows. Section 2 describes crucially about location estimation of a photo and its paradigms. The architecture of our G-sigMR is step-by-step explained in the section 3. The experiments and comparisons are in the section 4. For the conclusion, we summarize our G-sigMR workflow with its modern applications and the direction of our future work in the section 5.

2. Location Estimation of a Photo

Location estimation of a photo is the method to find the geographical location where the photo was taken. Since many million photos are shared by users on the social multimedia. Many photos are taken frequently from the same place. Some of them are either with or without geo-tagging. It is feasible that some geo-untagged photos can be estimated their locations with the help of million geo-tagged photos from the large-scale dataset. The researches about location estimation of a photo can be categorized into 2 paradigms: Visual-based and multisource-based evidence, respectively.

2.1. Visual-based Evidence

The photo representation (in term of visual contents) is considered to find the location. The first process known as feature detection is used to find the photo's characteristics (also known as key-points) that enable the photo to be distinguished from other photos [11]. The key-points of the same architectural scene or geography are not changed despite of the diversity of camera-viewpoints. However, feature detection produces only horizontal and vertical co-ordinates of the key-points from a photo [12] which are not enough for the landmark retrieval. It needs a descriptive vector (known as feature description) to describe the region surrounding the key-point [11-13] (also known as spatial information [14-17]). Feature descriptions of a photo are in term of a vector that represents the important visual contents of the architectural place or geography from a single photo [18-20]. Geo-tagged photos (in terms of vectors) and their locations are collected in the dataset [18-19] which later can be used in the landmark retrieval system for the indexing of many scenes with their locations [21-22]. For estimating the location of a geo-untagged photo, the most similar scene from the dataset is estimated that it was taken at the same location as the geo-untagged photo [23]. The most feature descriptions for the computation of vector similarity in the landmark retrieval are SIFT (Scale-invariant Feature Transform) [24-26], SURF (Speeded-Up Robust Features) [27] and HoG (Histogram of Gradients) [28-29]. However, those traditional feature descriptions are still having some deficiency in a vector representation of an architectural place or geography for landmark retrieval such as quantization loss, non-discrimination of vector [9, 14-15]. Most improved methods used the concept of visual synonym to be accomplished geometric coherence estimation [30-34]. A visual synonym is a pair of visual contents [35-38] that can be combined to find the location of a photo from the large-scale dataset by mapping to the similar scenes and reducing the irrelevance.

DOI:10.4186/ej.2017.21.3.295

2.2. Multisource-based Evidence

Location of a photo can be estimated using visual contents of the photo in combination with textual metadata which is called Multisource-based evidence. Textual metadata consists of annotations [39-41], tags [42-43], duplicated comments [44-47] and/or previous user's sharing [48] that are crawled from the social multimedia. The visual contents of a photo with its textual metadata are used to estimate the location of the photo [49-50]. For example, there is a coal-fire photo with many frequent comments about "Gate Hell" or "Door to Hell", the location (that the photo was taken) may be at Derweze, Turkmenistan. In contrast, multisource-based evidence exists totally deficiency. Since textual metadata are manually organized by million sharing on the social multimedia which affect to the correctness of textual information. If the frequent texts are not associated with the photo, it is easier to be higher errors in location estimation of a photo. Multisource-based evidence must have a thorough filtering which takes hugely more run-time than visual-based evidence. However, location estimation with the help of textual metadata will be work if the sufficiency of information on the social multimedia.

3. A Geo-signature MapReduce Workflow

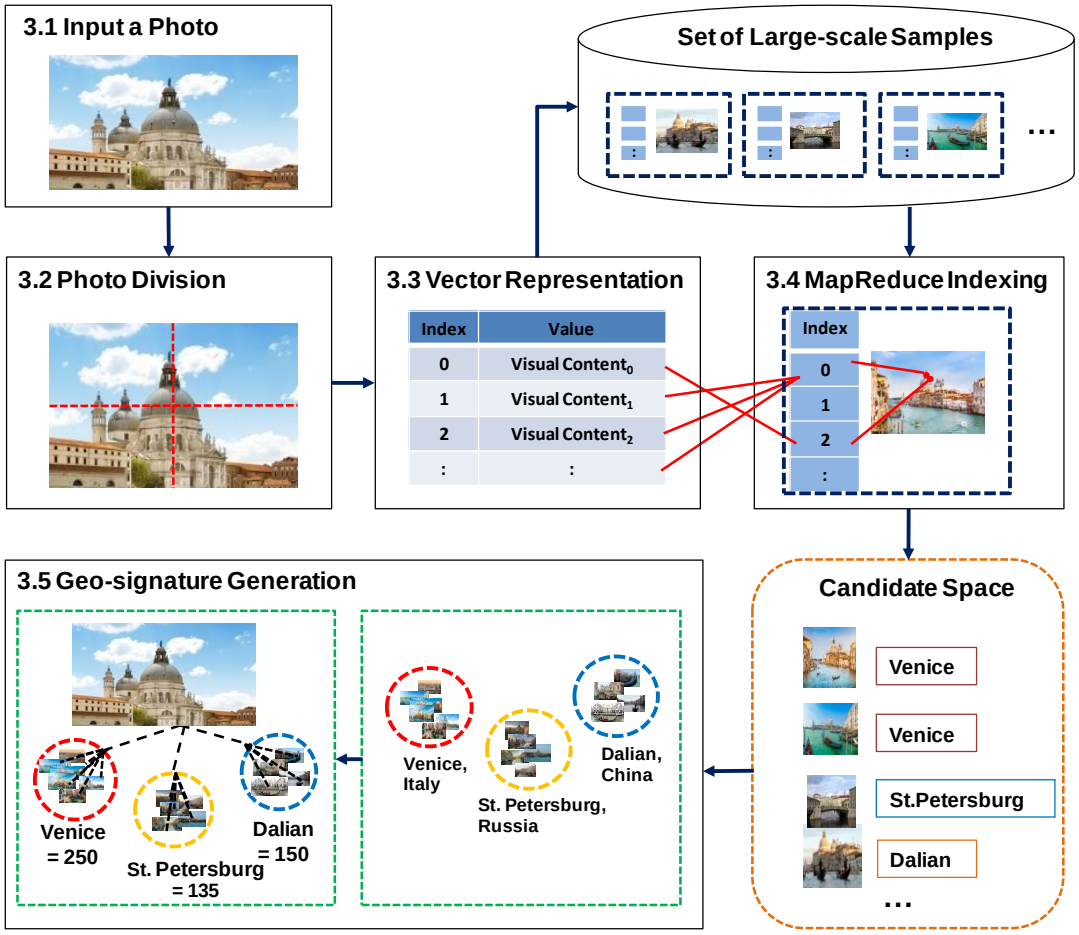

Fig. 2. The architecture of the G-sigMR workflow.

This section technically describes the architecture of Geo-signature MapReduce (G-sigMR) workflow which is a visual-based evidence as shown in Fig. 2. The G-sigMR can be organized into 5 main procedures: input a photo, photo division, vector representation, MapReduce indexing and geo-signature generation, respectively.

3.1. Input a Photo

Either geo-tagged or geo-untagged photo can be input to the workflow. In case of a geo-tagged photo, the photo with its geo-tagging will be incrementally added to the set of large-scale samples. In contrast, a geo-untagged photo will be estimated its location. The photo can be taken from any camera-viewpoint. Objects within the photo also can be located in any position as shown in Fig. 3.

Fig. 3. A geo-tagged or untagged photo input to the workflow.

3.2. Photo Division

The geo-tagged (or untagged) photo is divided into k parts. Gradients of the photo in the x-axis (G_x) and y-axis (G_y) are computed. Since the gradients are intensity of parts through the photo (also called gradient orientations) which are determined from the high-contrast visual contents such as edges or corners of architectural-objects within the photo. Range of gradients can be mathematically formulated according to Gaussian distribution (-1, 0, 1) in both horizon (denoted as G_x=[-1 0 1]) and vertical (denoted as a transpose of horizon G_y=[-1 0 1]^T). And the photo is divided into k parts as shown in Fig. 4., where k=4.

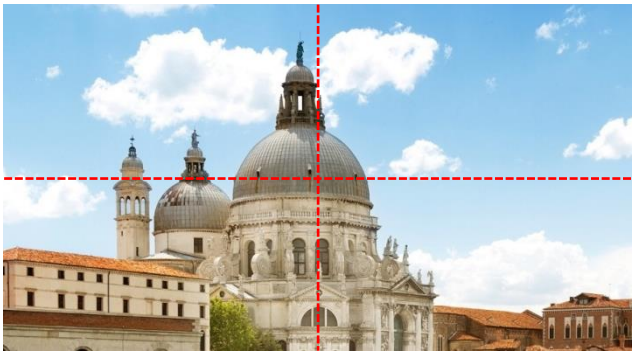

Fig. 4. A photo division into k parts (k=4)

After the photo division into k parts, the magnitude of gradient in each part ($G_{x,y}(i)$) is recursively computed by Eq. (1).

$$G_{x,y}(i) = \begin{cases} \phi & if \quad i > k \\ \sqrt{G_{x\ at\ i}^2 + G_{y\ at\ i}^2}, G_{x,y}(i+1) & if \quad i \le k \end{cases} \tag{1}$$

where i is a sequence of parts ($1 \le i \le k$), k is a number of parts within the photo, $G_{x\ at\ i}$ is a magnitude of gradient in the x-axis of the i-th part, $G_{y\ at\ i}$ is a magnitude of gradient in the y-axis of the i-th part and $G_{x,y}(i+1)$ is a magnitude of gradient of the next part (the $(i+1)$-th part).

The gradient orientation in each part ($\theta_{x,y}(i)$) is also computed in term of the slope of x and y axis by Eq. (2).

DOI:10.4186/ej.2017.21.3.295

$$\theta_{x,y}(i)=\begin{cases}\phi & \text{if } i>k\\ \tan^{-1}\left(\dfrac{G_{y\ at\ i}}{G_{x\ at\ i}}\right),\ \theta_{x,y}(i+1) & \text{if } i\le k\end{cases} \tag{2}$$

where i is a sequence of parts ($1 \le i \le k$), k is a number of parts within the photo, $G_{x\ at\ i}$ is a magnitude of gradient in the x-axis of the *i-th* part, $G_{y\ at\ i}$ is a magnitude of gradient in the y-axis of the *i-th* part and $\theta_{x,y}(i+1)$ is an orientation of gradient of the next part (the *(i+1)-th* part).

3.3. Vector Representation

Since the important (also called high-contrast) visual contents of a photo need a mathematical representation in term of a vector. From Fig. 5., each part from the previous procedure is divided into k sub-parts (k=4). In other words, the photo is further divided into k^2 sub-parts.

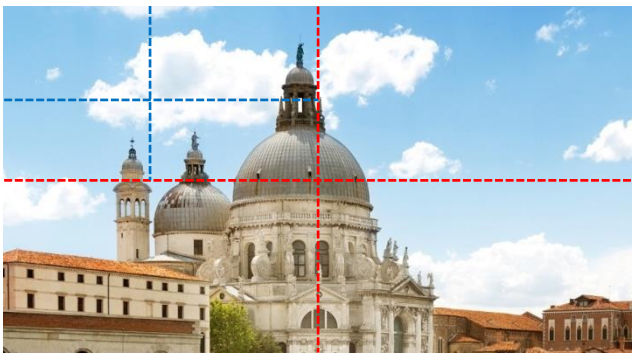

Fig. 5. Each part division into k sub-parts (known as k^2 sub-parts of the photo).

The intensity from pixels in each sub-part within a part is also recursively computed using the overlapping of local contrast normalization with the measure of intensity between parts to be robust for changeable in terms of illumination, scale, rotation and shadowing, as in a series of the orientation of gradient in each sub-part ($q_{x,y}(i,j)$) is recursively computed by Eq. (3).

$$q_{x,y}(i,j)=\begin{cases}\phi & \text{if } i>k\\ q_{x,y}(i+1,j) & \text{if } i\le k \text{ and } j>k^2\\ q_{x,y}(i,j+1) & \text{if } q_{x,y}(i,j)\notin B,\ i\le k \text{ and } j\le k^2\\ \tan^{-1}\left(\dfrac{C_{y\ at\ j}}{C_{x\ at\ j}}\right)+q_{x,y}(i,j+1) & \text{if } q_{x,y}(i,j)\in B,\ i\le k \text{ and } j\le k^2\end{cases} \tag{3}$$

where i is a sequence of parts ($1 \le i \le k$), k is a number of parts within the photo, j is a sequence of sub-parts ($1 \le j \le k^2$) within the *i-th* part, k^2 is a number of sub-parts within the photo, B is a set of important visual contents, $C_{x\ at\ j}$ is a gradient of the *j-th* sub-part in the x-axis, $C_{y\ at\ j}$ is a gradient of the *j-th* sub-part in the y-axis, $q_{x,y}(i,j+1)$ is a series of the direction of gradient in the *i-th* part and the next sub-part (the (j+1)-*th* sub-part) and $q_{x,y}(i+1,j)$ is a series of the orientation of gradient in next the part (the *(i+1)-th* part) and the *j-th* sub-part.

Only the important (or high-contrast) visual contents from the photo are represented in term of a vector as computed by Eq. (4) that considers from the magnitude of gradient in the *i-th* part ($G_{x,y}(i)$, from Eq. (1)), the orientation of gradient in the *i-th* part ($\theta_{x,y}(i)$, from Eq. (2)) and the series of orientation of gradient in the *j-th* sub-part ($q_{x,y}(i,j)$, from Eq. (3)), respectively.

$$represent(j)=\begin{cases}\phi & if \quad j>k^2\\ (G_{x,y}(i)*q_{x,y}(i,j))\left|\theta_{x,y}(i),\ represent(j+1)\right. & if \quad j\le k^2\end{cases} \tag{4}$$

where i is a sequence of parts ($1 \le i \le k$), j is a sequence of sub-parts ($1 \le j \le k^2$) within the i-th part, k^2 is a number of sub-parts within the photo and $represent(j+1)$ is a representation of important visual contents from the next sub-part (the $(j+1)$-th sub-part).

3.4. MapReduce Indexing

Customarily, the purpose of "MapReduce approach" [51-53] was designed for the actual "dynamic structure" for the "velocity", "volume" and "variety" of non-volatile large-scale datasets (or Big data [1]). In case of a geo-untagged photo, MapReduce indexing filters only the useful geo-tagged photos (that were collected in term of vectors with geo-tagging) from the set of large-scale samples which are similar to some visual contents of geo-untagged photo. The similarity between the geo-untagged photo and some geo-tagged photos from the set of large-scale samples is recursively computed using $index(P_k, P_*, a)$ as Eq. (5).

$$index(P_k,P_*,a)=\begin{cases}\phi & if \quad k>a\\ map(F_i,F_j,m,n),\ index(P_{k+1},P_*,a) & if \quad k\le a\end{cases} \tag{5}$$

where k is a sequence of geo-tagged photos from the dataset($1 \le k \le a$), a is a number of geo-tagged photos in the set of large-scale samples, P_* is the geo-untagged photo, P_k is the k-th geo-tagged photo from the set of large-scale samples and $index(P_{k+1}, P_*, a)$ is a similarity between the next geo-tagged photo (the $(k+1)$-th geo-tagged photo) from the set of large-scale samples and the geo-untagged photo.

For the one-by-one MapReduce structure, the function named $map(F_i, F_j, m, n)$ is used to match the similar contents between the geo-untagged photo and the k-th geo-tagged photo from the set of large-scale samples. Some geo-tagged photos that have similar visual contents with the geo-untagged photo are chosen to store in the candidate space. If a content F_i (from the geo-untagged photo) is not similar to a content F_j (from the k-th geo-tagged photo), the $reduce(F_i, F_j)$ function eliminates them before execution in the geo-signature computation (the next procedure). In case of unknown (or unseen) location of a geo-untagged photo, the system generates the label of this photo with its visual contents (instead of Geo-signature generation) which will be added to the set of large-scale samples for the next time MapReduce Indexing of this similar photo.

The one-by-one MapReduce structure can be mathematically described in Eq. (6).

$$map(F_i,F_j,m,n)=\begin{cases}\phi & if \quad i>m\\ map(F_{i+1},F_j,m,n) & if \quad j>n \quad and \ i\le m\\ reduce(F_i,F_j),map(F_i,F_{j+1},m,n) & if \ \overline{F_i\oplus F_j}=0,\ j\le n \ and \ i\le m\\ n(F_s)+1,map(F_i,F_{j+1},m,n) & if \ \overline{F_i\oplus F_j}=1,\ j\le n \ and \ i\le m\end{cases} \tag{6}$$

where i is a sequence of the i-th visual content of the geo-untagged photo, j is a sequence of the j-th visual content of the k-th geo-tagged photo from the set of large-scale samples, m is a number of visual contents within the geo-untagged photo, n is a number of visual contents within the k-th geo-tagged photo from the set of large-scale samples, F_i is the i-th visual content of the geo-untagged photo, F_j is the j-th visual content of the k-th geo-tagged photo from the set of large-scale samples, F_s is the visual content similarity between the geo-untagged photo and the k-th geo-tagged photo from the set of large-scale samples, $n(F_s)$ is a number of $F_s(s)$, $map(F_{i+1}, F_j, m, n)$ is matching between the next visual content (the $(i+1)$ –th visual content) of the geo-untagged photo and the j-th visual content of the k-th geo-tagged photo from the set of large-scale samples and $map(F_i, F_{j+1}, m, n)$ is matching between the i-th visual content of the geo-untagged photo and the next content (the $(j+1)$ –th visual content) of the k-th geo-tagged photo from the set of large-scale samples.

DOI:10.4186/ej.2017.21.3.295

3.5. Geo-signature Generation

Some geo-tagged photos from the set of large-scale samples have been chosen (from the previous MapReduce indexing) to store in the candidate space because their visual contents are similar with some contents of geo-untagged photo. Within the candidate space, this procedure groups the scenes that were taken from the same location into the same group as shown in Fig. 6.

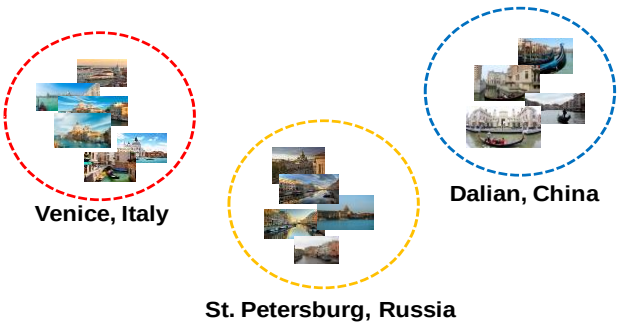

Fig. 6. Grouping of the geo-tagged photos from the candidate space that taken from the same location.

The uniqueness (known as unique evidences) of the *i-th* group named Geo-signature ($GeoSig_i$) is summarized from photos within the *i-th* group in term of visual synonyms that are associated with the important visual contents of geo-untagged photo. In other words, the Geo-signature of each group is generated from the similarity of geo-untagged photo which can be generated from Eq. (7).

$$Geo(G_i) = \begin{cases} \phi & if\ i > a \\ GeoSig_i = \sum_{j=1}^{m} n(F_s)_j,\ Geo(G_{i+1}) & if\ i \le a \end{cases} \tag{7}$$

where i is a sequence of the *i-th* group, j is a sequence of the *j-th* geo-tagged photo within the *i-th* group, m is a number of geo-tagged photos within the *i-th* group, a is a number of groups, G_i is the *i-th* group, F_s is the visual content similarity between the geo-untagged photo and the *k-th* geo-tagged photo from the set of large-scale samples, $n(F_s)$ is a number of $F_s(s)$, $GeoSig_i$ is a Geo-signature of the *i-th* group and $Geo(G_{i+1})$ is computing the Geo-signature of the next group (the *(i+1) -th* group).

Geo-signature of each group is used to find the location of geo-untagged photo. Since the geo-signature of each group is defined from the geo-untagged photo. All groups are compared together; resulting in a group has the highest unique evidences about the geo-untagged photo as shown in Fig. 7.

The location of geo-untagged photo is the *i-th* group that has the highest Geo-signature. From Fig. 8., Venice has the highest Geo-signature of the geo-untagged photo ($GeoSig_{Venice}$=250). Hence, the geo-untagged photo was taken from Venice, Italy which can be estimated by Eq. (8).

$$estimate(GeoSig_i, GeoSig_{i+1}) = \begin{cases} \phi & if\ \ i > a \\ estimate(\arg\max(GeoSig_i, GeoSig_{i+1}), GeoSig_{i+2}) & if\ \ i \le a \end{cases} \tag{8}$$

where i is a sequence of the *i-th* group, $GeoSig_i$ is a geo-signature of the *i-th* group, $GeoSig_{i+1}$ is a Geo-signature of the next group (the *(i+1) –th* group) and $GeoSig_{i+2}$ is a geo-signature of the next of next group (the *(i+2) –th* group).

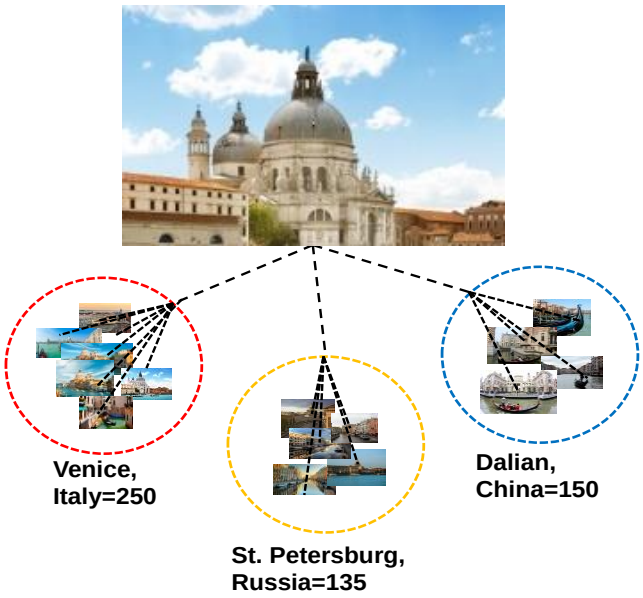

Fig. 7. Geo-signature of each group that generated from the geo-untagged photo.

4. Experiments and Comparisons

4.1. Experimental Design

In this section, the G-sigMR comparisons were made with other researches about location estimation of a photo: IM2GPS [3], SC [4], CS [5], VSA [6], MSER [7], and VCG [8], respectively. The photos used in our experiment were downloaded from GOLD (Geo-tagged large-scale dataset) [9-10] which contains more than 3.3 million geo-tagged photos. It covers more than 65,000 places around the world that partly focus on interesting places in China, Europe and America as shown in Fig. 8. GOLD was designed to be a test set for evaluation of location estimation and place recognition from a single photo (which can be directly downloaded from http://smiles.xjtu.edu.cn/Download/Download_gold.html).

Fig. 8. Some geo-tagged photos from GOLD [9-10].

The experimental setup was carried out under the environment of Intel Core(TM)2, Quad CP Q8400 and 48GB of RAM. Our G-sigMR was implemented by M-script language in Matlab R2015a as shown in Fig. 9. The source code can be requested for the performance comparison with our G-sigMR via the email (under in terms of use). For the experiment, the 80 places from GOLD were statistically used. Each location, 5,000 photos were randomly selected (as the same condition for evaluation in [8]).

DOI:10.4186/ej.2017.21.3.295

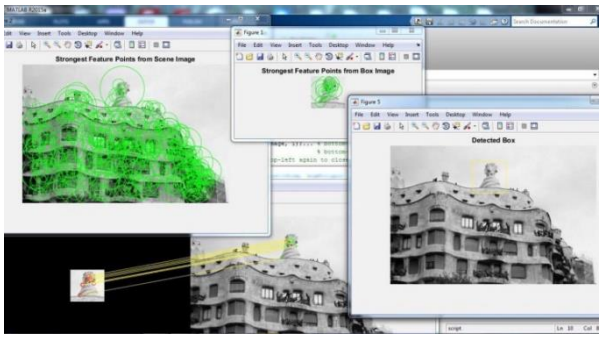

Fig. 9. Our G-sigMR implementation using M-script of Matlab R2015a.

4.2. Criteria Evaluation

Since the randomly selected 5,000 photos for one location can be evaluated by a criteria named "recognition rate" of the *i-th* location (denoted as RR_i). The RR_i are computed using TP_i, TN_i, FP_i and FN_i. For each location, the RR_i was computed by their random selection of 5,000 photos using Eq. (9).

$$RR_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \times 100 \tag{9}$$

where i is a sequence of the *i-th* location ($1 \le i \le 80$ because of 80 places), TP_i is a number of photos taken from the *i-th* location and correctly estimated, TN_i is a number of photos not taken from the *i-th* location and correctly estimated, FP_i is a number of photos not taken from the *i-th* location but wrongly estimated, and FN_i is a number of photos taken from the *i-th* location but wrongly estimated.

In this paper, we used a criteria named "average recognition rate" ($AVG(RR_i)$) of 80 places as a criteria for G-sigMR comparison with other researches. The $AVG(RR_i)$ can be computed from Eq. (10), where L is number of locations which equals as 80.

$$AVG(RR_i) = \frac{\sum_{i=1}^{L} RR_i}{L} \tag{10}$$

The performance comparison between our G-sigMR and other methods for estimating the location based on 3.3 million geo-tagged photos of GOLD [9-10] is shown in Fig. 10.

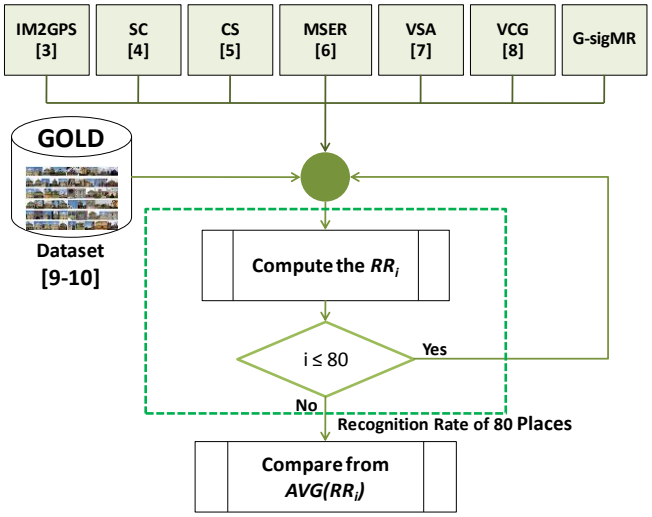

Fig. 10. Flow of performance evaluation and comparison.

4.3. Results and Discussions

The average recognition rate ($AVG(RR_i)$) of different methods are illustrated in Fig. 11. From the results, our G-sigMR produced more correctness in term of $AVG(RR_i)$ which equals as 94.17.

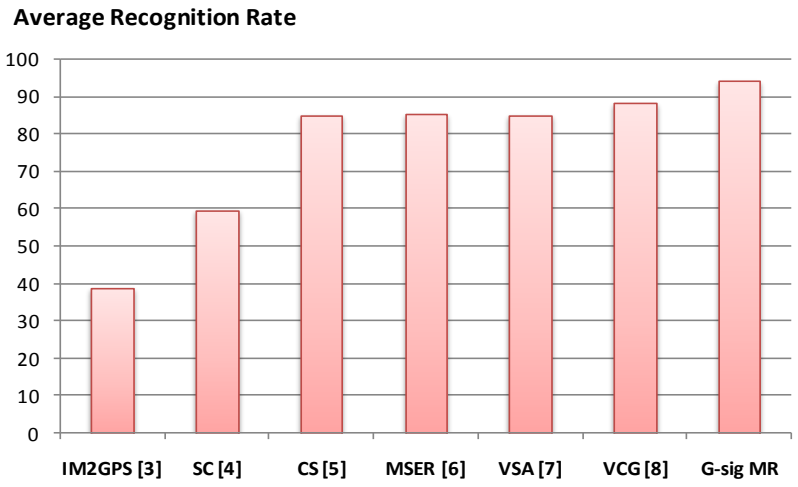

Fig. 11. The $AVG(RR_i)$ of IM2GPS [3], SC [4], CS [5], MSER [6], VSA [7], VCG [8] and our G-sigMR using GOLD [9-10]

Traditional approaches that directly estimate the location from a photo such as IM2GPS [3] and SC [4] were lower than 60%. The visual synonym approaches are VSA [6] ($AVG(RR_i)$ = 85.01), MSER [7] ($AVG(RR_i)$ = 85.47) and VCG [8] ($AVG(RR_i)$ = 88.16) which obviously can help for improvement of the recognition performance. However, they need a summarization of unique evidence from one place that to be distinguished from other places. Moreover, the structure of indexing is also useful for location estimation of a photo (as CS [5]). The G-sigMR is able to correctly estimate more than 90%. Since the workflow defines the uniqueness of a location named Geo-signature based on the visual synonym. Its Map-Reduce structure is used to index the similar scenes from million geo-tagged photos from the set of large-scale samples.

5. Conclusion

In this paper, we have designed the G-sigMR (Geo-signature MapReduce) workflow for location estimation of a photo. The previous approaches are not enough for defining the uniqueness of a place. However, G-sigMR could generate the uniqueness named Geo-signature of a place with the MapReduce structure suitable for large-scale dataset. In the experiment, we used 3.3 million geo-tagged photos from the Geo-tagged large-scale dataset (GOLD) which covers more than 65,000 places in China, Europe and America. G-sigMR is implemented using M-script in Matlab 2015a. For the evaluation, G-sigMR is also compared to other approaches: IM2GPS, SC, CS, MSER, VSA and VCG. In each test, the 80 places are statistically used. Each location is randomly selected 5,000 photos. From the results, G-sigMR produces a better performance in term of average recognition rate that is satisfactory (above 90%). The G-sigMR can be useful in modern applications, i.e., in social multimedia as a new function for location estimation from social photos; a search-based tourism application for finding the family photos of a place; and some forensic applications such as criminal location investigation from the scene for polices [54].

For future work, the direction of our G-sigMR will combine with the textual metadata such as annotations, tags, duplicated comments, and users' post from the social multimedia. Textual metadata will be filtered and summarized as the unique words or phrases of the geographical location by deep learning. Since the photo location estimation with the help of textual metadata has the challenge about information correctness that arbitrarily shared from social users. Textual metadata will be work for photo location estimation if information available on the social multimedia covers sufficiently all places around the world.

References

[1] EMC Information Infrastructure, "The digital universe of opportunities: Rich data and the increasing value of the Internet of things," IDC, 2016.

[2] R. Ji, Y. Gao, W. Liu, X. Xie, Q. Tian, and X. Li, "When location meets social multimedia: A survey on vision-based recognition and mining for geo-social multimedia analytics," *ACM Transaction Intelligent System Technology*, vol. 6, no. 1, pp. 1-18, 2015.

[3] J. Hays and A. A. Efros, "IM2GPS: Estimating geographic information from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[4] W. Zhou Y. Lu, H. Li, Y. Song, and Q. Tian, "Spatial coding for large scale partial-duplicate web image search," in *ACM International Conference on Multimedia*, 2010, pp. 511-520.

[5] J. Li. X. Qian, Y. Y. Tang, L. Yang, and T. Mei, "GPS estimation for places of interest from social users' uploaded photos," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 2058-2071, 2013.

[6] Z. Wu, Q. Ke, M. Isard, and J. Su, "Bundling features for large scale partial-duplicate web image search," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 2009, pp. 25-32.

[7] O. A. B. Penatti, F. B. Silva, E. Valle, V. Gouet-Brunet, and R. D. S. Torres, "Visual word spatial arrangement for image retrieval and classification," *Pattern Recognition*, vol. 47, no. 2, pp. 705-720, 2014.

[8] X. Qian, Y. Zhao, and J. Han, "Image location estimation by salient region matching," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4348-4358, November 2015.

[9] J. Li, X. Qian, Y. Y. Tang, L. Yang, and C. Liu, "GPS estimation from users' photos," *Journal of Magnetism and Magnetic Materials*, no. 1, pp. 118-129, 2013.

[10] X. Qian, X. Tan, Y. Zhang, R. Hong, and M. Wang, "Enhancing sketch-based image retrieval by re-ranking and relevance feedback," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 195-208, January 2016.

[11] A. A. Olaode, G. Naghdy, and C. A. Todd, "Unsupervised classification of images: A review," *International Journal of Image Processing*, vol. 8, no.5, pp. 325-342, 2014.

[12] J. Areeyapinan, P. Kanongchaiyos, and A. Kawewong, "Using multi-descriptors for real time cosmetic image retrieval," *Engineering Journal*, vol. 18, no.4, pp. 97-111, 2014.

[13] L. Soimart and P. Mookdarsanit, "Gender estimation of a portrait: Asian facial-significance framework," in *The 6th International Conference on Sciences and Social Sciences*, 2016.

[14] A. A. Olaode, G. Naghdy, and C. A. Todd, "Efficient region of interest detection using blind image division," in *Signal Processing Symposium*, 2015, pp. 1-6.

[15] A. A. Olaode, G. Naghdy, and C. A. Todd, "Bag-of-visual words codebook development for the semantic content based annotation of images," in *IEEE 11th International Conference on Signal-Image Technology & Internet-Based Systems*, 2015, pp. 7-14.

[16] L. Soimart and M. Ketcham, "The segmentation of satellite image using transport mean-shift algorithm," in *The 13th International Conference on IT Applications and Management*, 2015, pp. 124-128.

[17] L. Soimart and M. Ketcham, "An efficient algorithm for earth surface interpretation from satellite imagery," *Engineering Journal*, vol. 20, no. 5, pp. 215-228, 2016.

[18] H. Debnath and C. Borcea, "TagPix: Automatic real-time landscape photo tagging for smartphones," in *IEEE Conference on MOBILe Wireless MiddleWARE*, Bologna, 2013, pp. 176-184.

[19] Y. Li, D. J. Crandall, and D. P. Huttenloche, "Landmark classification in large-scale image collections," in *IEEE Conference on Computer Vision*, Kyoto, 2009, pp. 1957-1964.

[20] G. Tolias, Y. Avrithis, and H. Jegou, "Image search with selective match kernels: Aggregation across single and multiple images," *International Journal of Computer Vision*, vol. 116, no. 3, pp. 247–261, 2015.

[21] D. M. Chen, G. Baatz, and K. Köser, "City-scale landmark identification on mobile devices," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 737-744.

[22] A. Torii, J. Sivic, M. Okutomi, and T. Pajdla, "Visual place recognition with repetitive structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2346-2359, 2013.

[23] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific fusion for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 803-815, 2012.

[24] D. G. Lowe, "Object recognition from local scale-invariant features," in *IEEE Conference on Computer Vision*, Kerkyra, 1999, pp. 1150-1157.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004, pp. 91-110.
[26] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Conference on Computer Vision and Pattern Recognition*, Washington, DC, 1997, pp. 1000–1006.
[27] H. Bay, T. Tuytelaars, and L. V. Gool, "Speeded up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no.3, pp. 346–359, 2008.
[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Conference on Computer Vision and Pattern Recognition*, San Diego, 2005, pp. 886-893.
[29] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *Journal Computer Vision and Image Understanding*, vol. 117, no. 7, pp. 790-806, 2013.
[30] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, "Descriptive visual words and visual phrases for image applications," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 75-84.
[31] E. Gavves, C. G. M. Snoek, and A. W. M. Smeulders, "Visual synonyms for landmark image retrieval," *Compututer Vision Image Understand*, vol. 116, no. 2, pp. 238-249, 2012.
[32] J. Chen, B. Feng, L. Zhu, P. Ding, and B. Xu, "Effective near-duplicate image retrieval with image-specific visual phrase selection," in *IEEE International Conference on Image Processing*, Orlando, 2012, pp. 1909-1912.
[33] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8.
[34] W. Tang, R. Cai, Z. Li, and L. Zhang, "Contextual synonym dictionary for visual object retrieval,"in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 503-512.
[35] P. Mookdarsanit, L. Soimart, M. Ketcham, and N. Hnoohom, "Detecting image forgery using XOR and determinant of pixels for image forensics," in *IEEE 11th International Conference on Signal-Image Technology & Internet-Based Systems*, Bangkok, 2015, pp. 613-616.
[36] L. Soimart and M. Ketcham, "Hybrid of pixel-based and region-based segmentation for geology exploration from multi-spectral remote sensing," in *The 11th International Symposium on Natural Language Processing*, 2016, p. 74.
[37] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1709-1721, June 2015.
[38] X. Yang, X. Qian, and T. Mei, "Learning salient visual word for scalable mobile image retrieval," *Pattern Recognition*, vol. 48, no. 10, pp. 3093-3101, 2015.
[39] A. Shimada, H. Nagahara, R. I. Taniguchi, and V. Charvillat, "Geolocation based image annotation," in A*sian Conference on Pattern Recognition*, Beijing, 2011, pp. 657-661.
[40] X. Qian, X. Liu, X. Ma, D. Lu, and C. Xu, "What is happening in the video?—Annotate video by sentence," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1746-1757, 2016.
[41] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660-3671, 2016.
[42] P. Kelm, S. Schmiedeke, and T. Sikora, "How spatial segmentation improves the multimodal geo-tagging," in *Proceeding of Media Eval Workshop*, Pisa, 2012, pp. 1–2.
[43] D. Lu, X. Liu, and X. Qian, "Tag-based image search by social re-ranking," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1628-1639, 2016.
[44] X. Li, C. G. M. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1310-1322, Nov. 2009.
[45] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews,"*IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1910-1921, 2016.
[46] G. Zhao, X. Qian, and C. Kang, "Service rating prediction by exploring social mobile users' geographic locations," *IEEE Transactions on Big Data*, vol. PP, no.99, pp.1-1, 2016.
[47] G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors," *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 496-506, 2016.

[48] C. Hauff and G.-J. Houben, "Placing images on the world map: A microblog-based enrichment approach," in *International Conference on Research and Development in Information Retrieval*, New York, 2012, pp. 691–700.
[49] P. Mookdarsanit and M. Ketcham, "Image Location Estimation of Well-known Places from Multi-source based Information," in *The 11th International Symposium on Natural Language Processing*, 2016.
[50] S. Jiang, X. Qian, T. Mei, and Y. Fu, "Personalized travel sequence recommendation on multi-source big social media," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 43-56, 2016.
[51] P. P. Dhekale and N. Jichkar, "Efficient data search using map reduce framework," *World Conference on Futuristic Trends in Research and Innovation for Social Welfare*, Coimbatore, 2016, pp. 1-4.
[52] P. Mookdarsanit and S. Gertphol, "Light-weight operation of a failover system for cloud computing," in *IEEE 5th International Conference on Knowledge and Smart Technology*, Chonburi, Thailand, 2013, pp. 42-46.
[53] X. Qin, B. Kelley, and M. Saedy, "A fast map-reduce algorithm for burst errors in big data cloud storage," in *IEEE 10th System of Systems Engineering Conference*, San Antonio, TX, 2015, pp. 398-403.
[54] B. Todd and G. Botelho. (2015). *Tweet hows 'El Chapo,' official thinks - but when and where still a mystery* [Online]. Available: http://edition.cnn.com/2015/09/07/americas/mexico-el-chapo-guzman-photo-tweet/ [Accessed: 12 December 2016]