Unsupervised and Fast Continent Classification of Digital Image Collections using Time

Frode Eika Sandnes

Faculty of Engineering
Oslo University College
P.O. Box 4 St. Olavs Plass 0130 Oslo, Norway

Abstract— **Advances in storage capacity means that digital cameras can store huge collections of digital photographs. Typically such images are given non-descriptive filenames names such as a unique identifier, often an integer. Consequently it is time-consuming and difficult to browse and retrieve images from large collections especially on small consumer electronics devices. A strategy for classifying images into geographical regions is presented which allows images to be coarsely sorted into the continent where they were taken. The strategy employs patterns in the time-stamps of images to identify events such as holiday and individual days, and to estimate the approximate longitude where the photographs were taken. Experimental evaluations demonstrate that the continent is correctly estimated for 89 % of the images in arbitrary collections and that the longitude is estimated with a mean error of 27.5 degrees. The strategy is relatively straightforward to implement, also in hardware, and computationally inexpensive.**

Keywords— **digital camera, mobile phone camera, geo-spatial tagging, data-mining, image browsing, image retrieval**

I. INTRODUCTION *(HEADING 1)*

Recent advances in digital camera and storage technology combined with falling costs has lead to an exploding increase in digital images and image collections. This has led to new challenges in terms of organizing, browsing and retrieving images in large image collections. One avenue of research is applying complex content based strategies where intelligent strategies are applied for determining the contents of images. More basic strategies simply exploit time, where images are classified according to named folders and presented chronologically. Sometimes, extra information is available [9, 11] such as the geographical location where the image was taken. Such geo-spatial information is usually acquired with a GPS device and can lead to different and more efficient ways of browsing images [1, 4]. However, GPS-based geo-tagging is problematic for several reasons. First, few digital cameras are equipped with GPS receivers. Second, GPS receivers usually need some time to lock onto the overhead satellites. Third, the GPS infrastructure is reaching the end of its lifetime [7].

Consequently, several strategies for extracting geographical information without GPS have been proposed. Digital sextants has been implemented using digital cameras for direct sun elevation observations [5]. A sun sensor based on direct sun observations has been proposed [14]. However, problems with direct sun observations are that the sun needs to be visible and the calculation of sun position is dependent on the optics used,

i.e., knowledge of the lens characteristics is needed. Moreover, the position of the sun needs to be compared to some other entity, usually the horizon. Efficient algorithms for horizon detection exist [3, 6], but it is not always feasible to detect the horizon.

Moreover, direct sun observations work best when the sun elevation is low as very wide angle optics is needed for large sun elevations close to 90 degrees. To combat the problems with dependence on optics and large sun elevations it has been proposed to instead determine the sun elevation indirectly from the length of shadows cast by objects [13]. However, such a strategy would require complex image processing and no automatic working system has yet been demonstrated. In a more feasible approach the exposure level of the scene where images were taken has been explored as a means for determining the approximate geographical position of images with a documented accuracy of about 15 degrees [12]. One drawback of this strategy is that it is dependent on meta information collected by the digital camera and embedded in the images using the EXIF-format [2]. Another strategy is to look at the intensity of the image content itself [8, 10]. However, this strategy has only been demonstrated for webcam images taken at static location at regular intervals. This approach is also computationally expensive. Alternative approaches involves identifying the location of images from the contents of images such as landmarks [15].

This study presents a strategy for classifying images into the continent of its origin, hence geo-localization of images with quite a rough precision. However, the proposed approach is simple to implement, computationally efficient as it does not consider image contents and it can be applied to images taken with modest digital camera hardware. Applications of the approach include the automatic geo-tagging of images for image collection organization, browsing and retrieval applications.

II. METHOD

The strategy proposed herein assumes that the digital camera is equipped with an internal clock (chronograph) that is set with the correct date and time accurate to the nearest hour, and that this clock remains constant for all the images to be classified. Most users do not bother to set the internal clock in their digital cameras and many do not even know how to set the time. Moreover, it is also assumed that the time zone of the clock is known, or that the universal time (UTC) is used. The

subsequent sections assume that the image timestamps are represented in universal time.

Next, the images are grouped into daily chunks, i.e., all images in the range from 0:00 to 23:59 on a given date are considered one day. Next, groups of consecutive days are combined into events as such images are likely to be contextually related.

Images rely on light and most sightseeing photographs are therefore taken during daylight, or during the time of the day when the photographer is awake. Most people follow such daily rhythms were we sleep at night when it is dark and stay awake during the day when it is light.
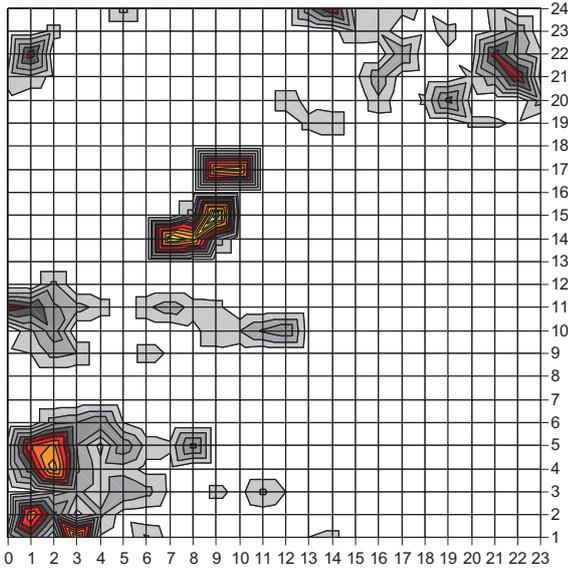

Figure 1.   Time stamp histograms plotted as a function of day.

Fig. 1 shows an example of the 24-hour histogram for the first events of the data used in this study. In this plot the horizontal axis represent histogram hour and the vertical axis represent day number. There are clear regions, or blobs, surrounding certain peaks. Clearly, at day 4 there is a peak around 2:00 UTC suggesting a Asia-Pacific set of images, at day 14 there is a peak running from around 7:00-9:00, and at day 17 a peak running from 9:00-10:00 UTC, both suggesting two sets of photographs taken in Europe or Africa, while, at day 23 there is a peak at 23:00 UTC suggesting a collection of photographs taken in the Americas.

A 24-bin histogram h with one bin for each hour of the day is constructed for each day where the number of photographs taken during each hour is counted. This histogram is shifted such that the mode, or the maximum bin is centered at the 12th, or middle bin, i.e., $h'_i = h_{12 - \bmod e}$ for all i's. The mean photograph time is then calculated as follows:

$$\bar{t} = \sum_{i=0}^{23} i \cdot h_i + \bmod e \tag{1}$$

The mean photograph universal time is a rough estimate of midday where the sun is at its maximum. Alternatively, the mode or the median could be used. Midday will therefore occur at the given universal time. The longitude of this location is therefore:

$$\lambda = \frac{360}{24}(12 - \bar{t}) \tag{2}$$

Where negative values signals degrees east and positive values signals degrees west.

In addition to an approximate longitudinal estimate a linguistic classification can also be computed. For this purpose the continent membership function $c(t)$ is introduced which gives the probability of the image taken at a particular time being in that continent. Fig. 2 shows the continent member functions for Antarctica, North America, South America, Africa, Europe, Asia and the Pacific. These were obtained by summing the proportion of land belonging to the continents for 10 degree longitudinal slices. Antarctica is omitted since it overlaps with all the other membership functions and very few people have a chance to visit this place. Moreover, North and South America are combined in the Americas, Africa and Europe into Europe-Africa and finally Asia and the Pacific are combined into one Asia-Pacific membership function because of the large overlaps.
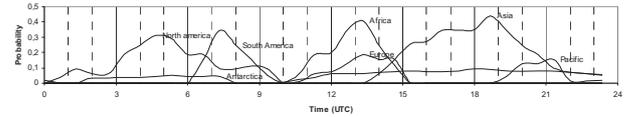

Figure 2.   The continent membership functions for the seven continents. Probability of continent membership is a function of longitude in universal time (UTC) measured in hours.

The continent membership functions for the seven continents. Probability of continent membership is a function of longitude in universal time (UTC) measured in hours.

$$S_{continent} = \sum_{i=0}^{23} V_{continent}(i) \tag{3}$$

and the vote function $V_{continent}(t)$ is given by

$$V_{continent}(t) = \begin{cases} h_t \ if \ m(t) > 0 \\ 0 \ if \ m(t) = 0 \end{cases} \tag{4}$$

After the score is computed for each continent, namely $S_{Americas}$, $S_{Europe-Africa}$ and $S_{Asia-Pacific}$ the one with largest score is the winner.

For an event comprising multiple days one may compute the scores for each day and sum the wins for each continent. The one with overall most wins for the events is the winner. If there is a tie between continents the days are classified according to their daily classification.

The strategy presented herein is computationally efficient as no image analysis is performed. First, the N images in the

collection must be scanned and segmented, which both have time complexity of $O(N)$. Next, the D days making up the entire collection is processed in D steps. For each step the histogram of that day is built, in the worst case N images giving time complexity of $O(D \cdot N)$, the median is computed which has constant time complexity $O(1)$ as the histogram as 24 bins. Finally, voting is performed for the D days with a time complexity of $O(D)$. Consequently, the strategy has a computational worst case complexity of $O(D \cdot N)$.

III. EXPERIMENTAL EVALUATION

The strategy presented herein was implemented in Java and was run on the author's personal image collection comprising 7,672 images taken at various locations around the world from the spring of 2005 to the spring of 2009 with a Sony DSC-F828. The images were manually grouped into location and situation and labeled for reference. The attributes of this test suite comprising 22 events are listed in Table I, including location, longitude, start-time of the series, number of days spanned by each event and the number of images in each event.

TABLE I.   EXPERIMENTAL TEST SUITE

event	location	longitude		start-time			Day.	Img.
1	Tainan	121°	East	28	Feb	2005	6	615
2	Tokyo	139°	East	13	Mar	2005	5	283
3	Yummy	121°	East	14	Apr	2005	2	203
4	Neihu	121°	East	31	May	2005	1	94
5	S. Illinois	89°	West	22	Jun	2005	9	768
6	Beihai	116°	East	30	Nov	2005	3	164
7	Hanoi	105°	East	12	Dec	2005	2	37
8	Tokyo	139°	East	16	Apr	2006	2	220
9	Tokyo	139°	East	23	Apr	2006	1	3
10	San Juan	66°	West	21	Jul	2006	8	780
11	Wuhan	108°	East	28	Aug	2006	1	7
12	Wuhan	108°	East	1	Sep	2006	5	358
13	Seoul	127°	East	24	Apr	2007	4	256
14	Oslo	10°	East	27	Sep	2007	1	268
15	Victoria	123°	West	8	Oct	2007	6	668
16	Oslo	10°	East	19	Nov	2007	2	339
17	Kaohsuing	121°	East	5	Feb	2008	3	290
18	Oslo	10°	East	23	Jun	2008	3	387
19	Paris	2°	East	10	Aug	2008	4	290
20	Beijing	116°	East	26	Oct	2008	5	803
21	Cape Town	18°	East	21	Feb	2009	6	641
22	Brisbane	153°	East	8	Jul	2009	2	198

Table II shows the results of the longitudinal estimations. The table lists the mean distance (error) between the estimated and actual longitude, the distance between the minimum observation and actual longitude, the distance between the maximum observation and actual longitude and linguistic classification of continent. Three regions are considered, namely the Americas, Asia-Pacific and Europe-Africa.

TABLE II.   EXPERIMENTAL RESULTS

Event	mean error	min error	max error	Linguistic classification
1	10°	91°	106°	Asia-Pacific
2	30°	41°	124°	Asia-Pacific
3	63°	46°	91°	Asia-Pacific
4	84°	76°	91°	Europe-Africa*
5	17°	89°	16°	Americas
6	31°	49°	101°	Asia-Pacific
7	5°	15°	15°	Asia-Pacific
8	86°	41°	124°	Asia-Pacific
9	64°	64°	64°	Americas*
10	16°	6°	84°	Americas
11	53°	42°	57°	Asia-Pacific
12	3°	27°	93°	Asia-Pacific
13	9°	38°	112°	Asia-Pacific
14	1°	65°	65°	Europe-Africa
15	20°	123°	42°	Americas
16	13°	50°	125°	Americas*
17	28°	14°	106°	Asia-Pacific
18	1°	95°	80°	Americas*
19	19°	43°	88°	Americas*
20	12°	86°	116°	Asia-Pacific
21	10°	102°	87°	Europe-Africa
22	30°	12°	78°	Asia-Pacific

A. Discussion

The results in Table II confirm that the proposed strategy successfully managed to classify the image collections into the correct continents for 17 of the 22 events, yielding success rate of 77 % in terms of events. In terms of images the success rate is 85.5 % since as 6.559 of the total of 7.672 images where classified into the correct continent. The reason the success rate is higher when considering individual images is that events with more images are more likely to be successfully classified than events comprising fewer images. The set of incorrectly classified events comprise fewer images totally.

In the image collection investigated events 4, 9, 16, 18 and 19 were incorrectly classified. Here, event 4, which comprised images taken in Neihu, Taiwan were classified as being in Europe. The longitudinal error was also large for this set, namely 84 degrees. An inspection of the image collection reveals the problem, namely that the images are all taken during a relatively short time interval of two hours during an evening walk around Neihu. Since these images were all taken at night the midday estimate was shifted and consequently matches that of the US, since it is midday in the US when it is evening in Taiwan, broadly speaking.

Similar arguments apply for the other erroneous cases such as event 9 where images taken in Tokyo were classified as belonging in the Americas, event 16 where images taken in Oslo were misclassified as belonging in the Americas and events 18 and 19 representing images taken in Oslo and Paris were misclassified as being in the Americas.

Longitudinal estimates based on the mean, median and mode time of the images taken for 24 hour bins are shown in Fig. 3. The results show that the mean time gave the overall most accurate result of 27.5 degrees. The median was slightly worse with a longitudinal accuracy of 32.7 degrees while the modes gave the least accurate estimate of the three measures of centrality, namely 46.7 degrees.
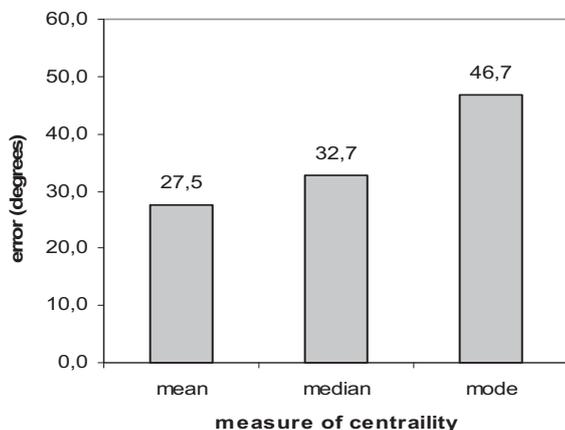

Figure 3.   Centrality measure of longitude.

B. Improving accuracy with voting

Table III shows an example of how the linguistic continent classification was improved with voting. The event represents images taken during a one-week visit to South Africa during the last week of February, 2009. The example shows that the first, third and sixth days were successfully classified as being in Europe-Africa, while the second day was misclassified as being in the Americas and fourth and fifth day were misclassified as being in Asia-Pacific. However, Europe-Africa won by the number of day-based voted for the event and the entire event was then successfully classified as images from Europe-Africa.

TABLE III.   VOTING FOR IMPROVING LINGUISTIC CLASSIFICATION

Day	mean longitude		no images	Americas	Europe -Africa	Asia -Pacific
1	12°	West,	144	54	58	8
2	7°	East,	211	114	28	77
3	29°	East,	67	8	46	13
4	5°	East,	84	5	23	24
5	73°	East,	19	0	2	19
6	9°	East,	111	3	67	20
			wins	1	3	2

This strategy is robust to events comprising days with images taken at nighttime. The number of nighttime images does not affect the results because the voting is day-based. If one considers the entire event then single days with many nighttime photographs may strongly bias the results negatively.
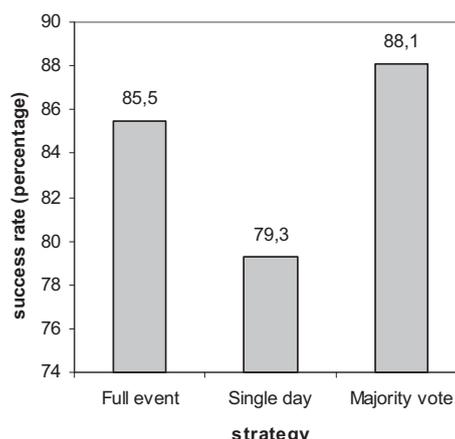

Figure 4.   Improving success rate with voting.

The effectiveness of the voting scheme is illustrated in Fig. 4. This illustration shows the success rate obtained with linguistic continent classification, the success rate obtained classifying according to single days and day-based majority vote for the entire event. Clearly, the majority vote yielded the highest success rate of 88.1 % while the event based strategy only achieved 85.5 %. Simply, considering individual days yielded a success rate of just 79.3 %.

Note that the proposed voting scheme may result in ties such as the one illustrated in Table IV. This example shows images taken in Paris, France during the second week of August, 2008. Clearly, the first and the last days are successfully classified as Europe-Africa, while the two middle days are incorrectly classified as America. If one had considered this as one event then the result would have been an overall misclassification of the Americas. Instead, the result is a tie between the Americas and Europe-Africa which is more correct. Note also that the numbers in each row of Tables III and IV do not tally – this is because images have membership in several continents due to the overlap in the membership function.

TABLE IV.   TIES ARE BETTER THAN MISCLASSIFICATION

day	Mean longitude		no. Images	Americas	Europe -Africa	Asia -Pacific
1	5°	West,	65	13	36	0
2	34°	West,	133	74	49	1
3	20°	West,	78	52	25	0
4	26°	East,	11	0	11	0
			Wins	2	2	0

IV.   LIMITATIONS

The strategy presented herein is based on three key assumptions. First, the images are time-stamped. This is a realistic assumption with most current day digital cameras that are equipped with an internal clock and calendar. Second, the time must be relatively correct and the relative offset from

UTC must be known. This is also a reasonable assumption as most camera owners set their camera to their home time-zone once they purchase the camera and never bother to change the time and date settings later. The strategy will thus not work for photographers who tamper with the date and time settings of their cameras. Third, the strategy assumes that the images are taken during daytime or spread throughout the day. This assumption is based on the fact that photography relies on light and the conditions for good photography at night is not as good. Clearly, there are many exceptions to this. As confirmed by the results herein, the proposed strategy does not work with images taken at narrow time-intervals at night. Moreover, this study does not address the problem of image collections comprising images taken with multiple cameras.

V. CONCLUSIONS

A simple strategy for effectively determining where digital images were taken was proposed. Time was used to estimate the approximate local midday for photographic events in image collections. The local midday then directly translates into longitude. A simple voting scheme was employed to realize a robust continent classifier. The experimental results demonstrate that the continent classifier achieves a success rate of 88.1%. Moreover, the longitude estimates based on 24-hour time means yield a longitudinal accuracy of 27 degrees. The proposed approach is computationally efficient and simple. It does not rely on image contents or complex meta-information as it only relies on consistently time-stamped images. It can for instance be easily integrated in low-cost digital cameras for unsupervised approximate geo-tagging as most such cameras already have an internal clock and calendar.

REFERENCES

[1] Ahern, S., Naaman, M., Nair, R., and Hui-I Yang, J., "World explorer: visualizing aggregate data from unstructured text in geo-referenced collections," in the proceedings of 7th ACM/IEEE-CS joint conference on Digital libraries, 2007, pp. 1-10.

[2] Alvarez, P., "Using Extended File Information (EXIF) File headers in Digital Evidence Analysis," International Journal of Digital Evidence, 2, 3, 2004

[3] Bao, G.-Q., Xiong, S.-S., and Zhou, Z.-Y., "Vision-based horizon extraction for micro air vehicle flight control," IEEE Transactions on Instrumentation and Measurement, 54, 3 (2005), pp. 1067-1072.

[4] Carboni, D., Sanna, S., and Zanarini, P., "GeoPix: image retrieval on the geo web, from camera click to mouse click," in the proceedings of Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, 2006, pp. 169-172.

[5] Cozman, F. and Krotkov, E., "Robot localization using a computer vision sextant," in the proceedings of IEEE International Conference on Robotics and Automation, 1995, pp. 106-111.

[6] Ettinger, S. M., Nechyba, C., and lfju, P. G., "Towards Flights autonomy: Vision-based horizon detection for micro air vehicles," in the proceedings of IEEE International Conference on Robotics and Automation, 2002.

[7] GAO, "GLOBAL POSITIONING SYSTEM: Significant Challenges in Sustaining and Upgrading Widely Used Capabilities," United States Government Accountability Office 2009.

[8] Jacobs, N., Roman, N., and Pless, R., "Toward Fully Automatic Geo-Location and Geo-Orientation of Static Outdoor Cameras," in the proceedings of IEEE Workshop on Applications of Computer Vision, 2008, pp. 1-6.

[9] Jang, C.-J., Lee, J.-Y., Lee, J.-W., and Cho, H.-G., "Smart Management System for Digital Photographs using Temporal and Spatial Features with EXIF metadata," in the proceedings of 2nd International Conference on Digital Information Management, 2007, pp. 110-115.

[10] Ray, S. F., "Camera Exposure Determination," in The Manual of Photography: Photographic and Digital Imaging, R. E. Jacobson, S. F. Ray, G. G. Atteridge, and N. R. Axford, Eds.: Focal Press, 2000.

[11] Romero, N. L., Chornet, V. V. G. C. G., Cobos, J. S., Carot, A. A. S. C., Centellas, F. C., and Mendez, M. C., "Recovery of descriptive information in images from digital libraries by means of EXIF metadata," Library Hi Tech, 26, 2 (2008), pp. 302-315.

[12] Sandnes, F. E., "Geo-Spatial Tagging of Image Collections using Temporal Camera Usage Dynamics," in the proceedings of I-SPAN 2009, 2009, pp. 160-165.

[13] Sandnes, F. E., "Sorting holiday photos without a GPS: What can we expect from contents-based geo-spatial image tagging?," Lecture Notes on Computer Science, 5879, (2009), pp. 256-267.

[14] [Trebi-Ollennu, A., Huntsberger, T., Cheng, Y., and Baumgartner, E. T., "Design and analysis of a sun sensor for planetary rover absolute heading detection.," IEEE Transactions on Robotics and Automation, 17, 6 (2001), pp. 939 - 947.

[15] Zheng, Y.-T., Ming, Z., Yang, S., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., and Neven, H., "Tour the world: Building a web-scale landmark recognition engine," in the proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), 2009, pp. 1085 - 1092.