



OPEN ACCESS

EDITED BY

Maria Luisa Sapino,
University of Turin, Italy

REVIEWED BY

Klimis Ntalianis,
University of West Attica, Greece
Eftychios Protopapadakis,
National Technical University of Athens, Greece

*CORRESPONDENCE

Stamatios Giannoulakis
✉ s.giannoulakis@cut.ac.cy

RECEIVED 22 January 2023

ACCEPTED 06 June 2023

PUBLISHED 04 July 2023

CITATION

Giannoulakis S, Tsapatsoulis N and Djouvas C
(2023) Evaluating the use of Instagram images
color histograms and hashtags sets for
automatic image annotation.
Front. Big Data 6:1149523.
doi: 10.3389/fdata.2023.1149523

COPYRIGHT

© 2023 Giannoulakis, Tsapatsoulis and Djouvas.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Evaluating the use of Instagram images color histograms and hashtags sets for automatic image annotation

Stamatios Giannoulakis^{1*}, Nicolas Tsapatsoulis² and
Constantinos Djouvas¹

¹Department of Communication and Internet Studies, Cyprus University of Technology, Limassol, Cyprus, ²Department of Public Communication, Cyprus University of Technology, Limassol, Cyprus

Color similarity has been a key feature for content-based image retrieval by contemporary search engines, such as Google. In this study, we compare the visual content information of images, obtained through color histograms, with their corresponding hashtag sets in the case of Instagram posts. In previous studies, we had concluded that less than 25% of Instagram hashtags are related to the actual visual content of the image they accompany. Thus, the use of Instagram images' corresponding hashtags for automatic image annotation is questionable. In this study, we are answering this question through the computational comparison of images' low-level characteristics with the semantic and syntactic information of their corresponding hashtags. The main conclusion of our study on 26 different subjects (concepts) is that color histograms and filtered hashtag sets, although related, should be better seen as a complementary source for image retrieval and automatic image annotation.

KEYWORDS

Instagram hashtags, Instagram images, histogram, Bhattacharyya distance, word embedding, automatic image annotation

1. Introduction

Nowadays, online content is generated in an unprecedented pace, including the publication of images on different platforms and fora. Of a paramount importance, however, is the use of efficient and effective techniques that allows the accurate retrieval of images. Image retrieval methods fall within three main categories: text-based image retrieval (TBIR), content-based image retrieval (CBIR), and automatic image annotation (AIA). The text-based techniques are inspired by document retrieval; Keywords are associated with images, e.g., the name of the image file, which are then used as text-based elements against which users' keywords will be matched. CBIR; on the contrary, images are retrieved according to their visual content. Given an example (target) image, CBIR transforms it to a feature vector, which is then used for retrieving images using a similarity metric among their feature vectors. Finally, the AIA approach has as a main idea to automatically learn semantic concept models from a large number of image samples and use these concepts to automatically annotate new images, i.e., assign labels to them. Therefore, AIA can be considered as a combination of TBIR and CBIR because it uses both text-based annotation and content-based image features. AIA is also a method that bridges the semantic gap between low-level image features and high-level semantics (Zhang et al., 2012).

Usually, images are indexed by their visual content based on low-level characteristics, such as color, texture, shape, and spatial layout (Latif et al., 2019). In practice, color as low-level feature is used for image classification and matching because of its effective and low computational cost (Chen et al., 2020). Among them, color histograms (Sergýán, 2008) are quite popular (Takeishi et al., 2018). An advance of color histograms is that are invariant to orientation and scale, and this feature makes it more powerful in image classification. Evidence to the above is the plethora of research studies using color histograms for image retrieval (Liu and Yang, 2013; Theodosiou, 2014; Mufarroha et al., 2020; Zhang et al., 2020).

The previous discussion clearly indicates the importance of efficient and accurate estimation of color-based similarity. In the case of automatic image annotation (AIA), however, images should be annotated with relevant labels. Different approaches exist in the literature for AIA (Tsapatsoulis, 2016, 2020). In this work, we are particularly interested in Instagram hashtags and the possibility of using them for AIA. This is because hashtags have some unique characteristics. They represent a specific topic, or idea, or annotation of an image or text, and they are used regardless of the topic they annotate (Kim and Seo, 2020). In addition, according to Gomez et al. (2020), hashtags is a form of image tagging. Furthermore, they have a metacommunicative use (Daer et al., 2014); The metacommunicative function of hashtags falls into five categories: emphasizing, iterating, critiquing, identifying, and rallying.

Despite their great potentials, only 25% Instagram hashtags are related to the visual content of Instagram images (Giannoulakis and Tsapatsoulis, 2015, 2016b). To alleviate this problem, in our previous study, we proposed different hashtags filtering techniques using classic (Giannoulakis and Tsapatsoulis, 2016a) and sophisticated (Giannoulakis et al., 2017; Giannoulakis and Tsapatsoulis, 2019) methods.

In this study, we attempt to bridge the semantic gap between image low-level features, such as color histogram and high-level semantic content. To do so, we investigated if AIA can be achieved with the aid of Instagram posts, assuming that Instagram is a rich source of implicitly annotated images. More precisely, we evaluate the coherency among similar images' low-level features and their corresponding filtered hashtags, i.e., we assume that similar images (e.g., images containing dogs) should have conceptually similar hashtags. To the best of our knowledge, this is the first research that quantifies the similarity of color histogram and hashtag in Instagram. We show that color histograms and filtered hashtag sets although related should be better seen as complementary source for image retrieval and automatic image annotation.

2. Image similarity and Bhattacharyya distance

Image similarity can be calculated using different approaches. Those can be broadly divided into two categories on the basis of the metrics that are used. Intensity-based approaches are based on features (indices) derived from pixel color intensities, while geometry-based approaches use geometric transformations between corresponding pixels (Deza and Deza, 2009; Li and Qian,

2015). In intensity-based similarity computation, the metrics that are usually used are correlation, chi-square, intersection, and Bhattacharyya distance (Arai, 2019; Forero et al., 2019). The geometry-based similarity metrics include pixel correspondence metric, closest distance metric, figure of merit, and partial Hausdorff distance metric (Prieto and Allen, 2003; Li and Qian, 2015). The main drawback of geometry-based similarity metrics is their high computational cost. Thus, intensity-based metrics, usually involving histogram and histogram matching, are frequently adopted.

While a variety of metrics is used for histogram matching, the most common metric is Bhattacharyya (Arai, 2019; Forero et al., 2019) distance, something that is also adopted in this study. Its basic principle is to calculate the distance between two probability distributions $p(x)$ and $q(x)$ which are approximated by the corresponding normalized (Equation 3) histogram vectors \vec{c}_1 and \vec{c}_2 (Zhang, 2006; Kayhan and Fekri-Ershad, 2021). Thus, the Bhattacharyya distance between two images I_1 and I_2 with histogram elements $c_1(x)$ and $c_2(x)$ were computed on a set of color hues \mathcal{X} as follows:

$$d(I_1, I_2) = -\ln(BC(\vec{c}_1, \vec{c}_2)), \quad (1)$$

$$BC(\vec{c}_1, \vec{c}_2) = \sum_{x \in \mathcal{X}} \sqrt{c_1(x) \cdot c_2(x)}, \quad (2)$$

where:

$$\sum_{x \in \mathcal{X}} c_1(x) = 1 \quad \sum_{x \in \mathcal{X}} c_2(x) = 1. \quad (3)$$

In order to use Bhattacharyya distance as a similarity metric, the reformulation indicated in Equation (4) was applied. It goes without saying, however, that different reformulations expressing a similar logic can be applied.

$$S(I_1, I_2) = \frac{1}{1 + d(I_1, I_2)} \quad (4)$$

It is clear from Equation (4), that the similarity among two image ranges in the $(0, 1]$ interval, with values close to 0 indicating very low similarity, while close to 1 denotes very high similarity (Han, 2015).

Bhattacharyya distance was widely used for computing the low-level content similarity of two images, video frames, or image regions, in a variety of purposes. Chacon-Quesada and Siles-Canales (2017) adopted Bhattacharyya distance as a metric for shot classification of soccer videos. In their effort to develop a moving target tracking algorithm, Ong et al. (2019) used color histograms of frame regions to locate the target object in each frame. Abidi et al. (2017) used histogram of oriented gradients (HoGs) and minimized the Bhattacharyya distance between two sets of gradient orientations expressing the desired and current camera poses, in their vision-based robot control system. Doulah and Sazonov (2017) clustered food-related images using Bhattacharyya similarity. The images were extracted from meal video captured with a wearable camera, and they were indexed using histograms in the HSV color space.

3. Word embeddings and Instagram hashtags

In the previous section, we discussed the distance between two histograms as a way to calculate the similarity of two images. In this section, we discuss the second key technology in our study, the word embeddings. Word embeddings were used for finding the similarity of hashtags, focusing on their use in the context of Instagram hashtags.

Word embeddings, i.e., techniques that convert words to numerical vectors retaining semantic and syntactic information, is a state-of-the-art approach in natural language processing, especially in document classification, sentiment analysis (Tsapatsoulis and Djouvas, 2019), and topic modeling (Argyrou et al., 2018; Tsapatsoulis et al., 2022). Word embedding techniques learn the relation between words via training on context examples of each word (Ganguly, 2020) using deep learning methods. Some of the most commonly used word embeddings are GloVe (Gomez et al., 2018), Word2vec (Jiang et al., 2020), and WordRank (Zhang, 2019). Pre-trained word embeddings in a variety of languages are available online, something that boosted their application on an impressive number of different fields.

Weston et al. (2014) used a convolutional neural network to create specific word embeddings for hashtags. The overall aim was to predict hashtags from the text of an Instagram post. Liu and Jansson (2017) tried to identify city events from Instagram posts and hashtags. They used Word2vec embeddings for query expansion, i.e., to identify terms related to the seed posts they used. Hammar et al. (2018) classified Instagram posts text into clothing categories using word embeddings. They used similarity matching via word embeddings to map text to their predefined ontology terms.

Prabowo and Purwarianti (2017) developed a system that helps online shop owners to response to Instagram comments. The system classifies the comments to those that are necessary to answer, those that the online shop owner needs to read, and those to ignore. By comparing the performance of support vector machines (SVMs) and convolutional neural networks (CNNs), they concluded that the combination of word embeddings with CNN learning provides the best combination.

Akbar Septiandri and Wibisono (2017) used Word2vec to detect spam comments on Indonesian Instagram posts. They used the *fastText* library which allows the easy expansion of word matching to short-text (paragraph) matching. Serafimov et al. (2019) proposed hashtag recommendation for online posts using word to paragraph matching with the aid Word2vec vectors. Gomez et al. (2018) combined images and caption to learn the relations between images, words, and neighborhoods, based on Instagram posts related to the city of Barcelona. To achieve their goal they used pretrained Gensim Word2Vec models to discover words that users relate with Barcelona's neighborhoods. Xu et al. (2020) used word embeddings to locate relevant documents in an information filtering system. The researchers produce a topic model tool and trained on users' interest documents. Then, the topic model was applied on incoming documents for estimating the relevance of the new documents to the user.

In the current study, we used GloVe (Gomez et al., 2018) pre-trained word embeddings model. GloVe is trained on Google News articles and Wikipedia content using the following optimization criterion (Pennington et al., 2014):

$$J = \sum_{i,j=1}^V f(X_{ij})(w_i^T \tilde{w} + b_i + \tilde{b}_j - \log X_{ij})^2 \quad (5)$$

where $f(X_{ij})$ tabulates the number of times word j occurs in the context of word i , $w \in \mathbb{R}^d$ are word vectors, $\tilde{w} \in \mathbb{R}^d$ are separate context word vectors, V is the size of the vocabulary, and b_i is a bias for w_i .

4. Comparative review

A brief summary of a related study and a comparative review is presented in this section of the study. Zhang et al. (2019) calculated the similarity between brands via posts of brands' followers of Instagram. Image feature extracted using 50-layer ResNet and ImageNet and tags converted into a vector with the help of Word2vec and *fasttext*TM. To measure the similarity, they use Pearson correlation and histogram similarity. The study differs from our study due to the different purpose and methodology used. The aim of Zhang et al. was to develop a marketing tool and not to locate the relation between image and hashtags as in our research.

Liu et al. (2015) studied the image color and text similarity in a software application called CITY FEED, that was used for classifying crowd sourced feeds. Image similarity analysis is based on color histogram and computed with Bhattacharyya coefficient. To calculate text similarity, the WordNet algorithm was used. Liu et al.'s research is similar to our research because we also use histogram and Bhattacharyya. However, the nature of the data are totally different. In our research, the data are Instagram photos belonging to a subject/hashtag (e.g., #dog) but are heterogeneous. Data from CITY FEED are not so heterogeneous due to the fact that photos are from the Municipality of Pavia.

5. Methodology

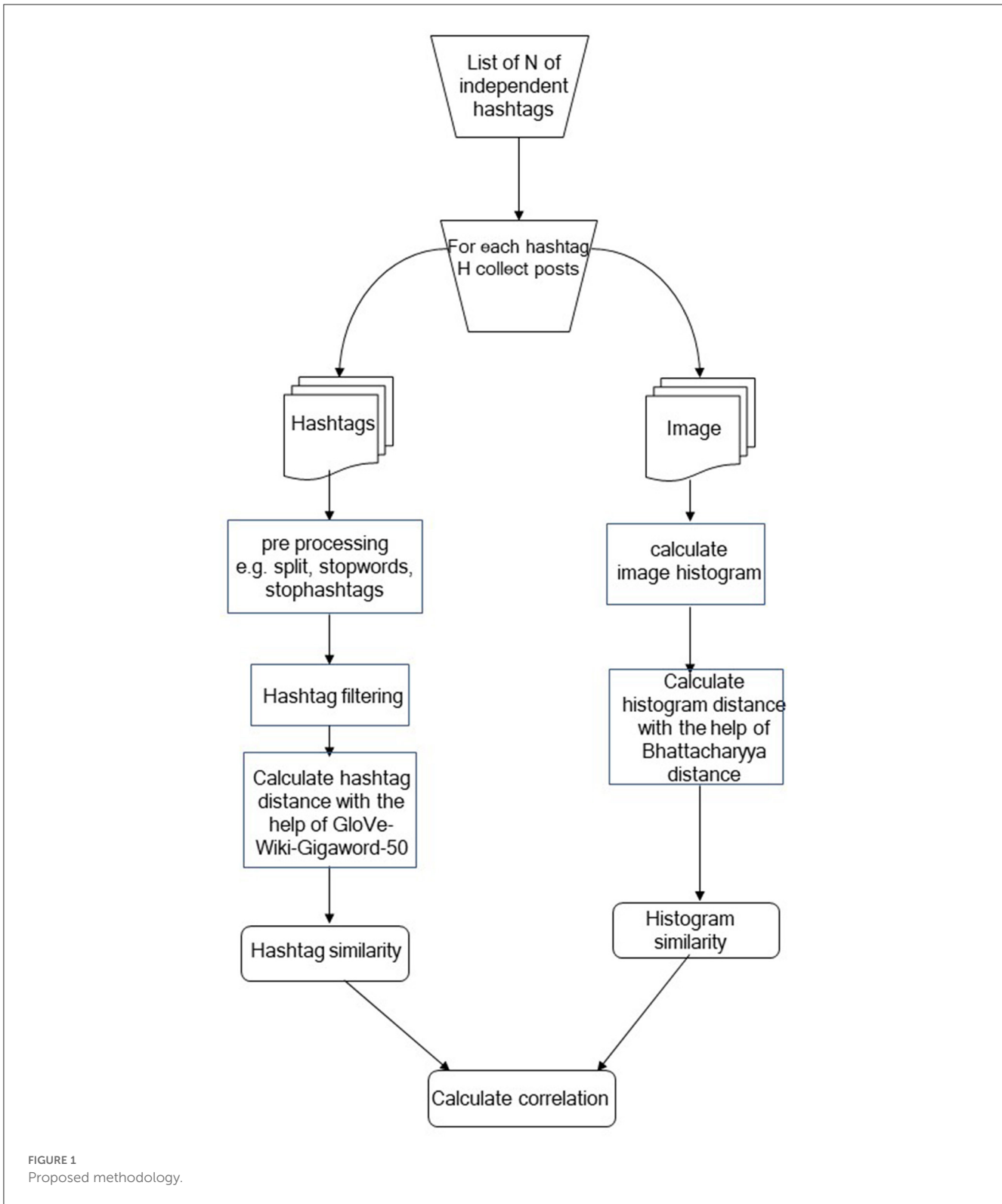
The current study was formulated as an experimental study expressed through two null hypotheses using two groups of Instagram posts, the relevant and the irrelevant subset.

H0₁: In relevant Instagram posts, there is no significant correlation between the similarity of the color histogram of image pairs and the similarity of their corresponding (filtered) hashtags sets.

H0₂: There is no significant difference in the average correlation between color histograms and hashtag sets in relevant and irrelevant posts.

Relevant Instagram posts are posts whose image match the hashtag subject and irrelevant those that their image do not match the hashtag subject (more details about the corpus used in Section 5.2).

At the same time, we expect that the correlation between color histograms and hashtag sets in relevant posts should be significantly higher than that of irrelevant posts.



In order to confirm or reject the null hypotheses, the process shown in Figure 1 was followed. First, we selected N independent hashtags, which in the context of the current study are referred to as hashtag subjects. For each hashtag subject, we searched Instagram creating two collections of Instagram posts, one containing posts whose image is visually relevant to the subject and one containing

posts whose image is not visually relevant to the subject. For posts collected using a hashtag subject, their corresponding image and hashtags were automatically collected using the Beautiful Soup¹ library of Python. For instance, if we searched Instagram using the

1 <http://www.crummy.com/software/BeautifulSoup/bs4/doc/>

TABLE 1 Average Bhattacharyya similarity scores (relevant posts).

	Bear	Cat	Chair	Dog	Dress	Elephant	Fish	Giraffe	Guitar	Hamster	Hat	Headband	Hedgehog
Mean	0.722	0.734	0.735	0.749	0.729	0.765	0.668	0.729	0.674	0.744	0.683	0.701	0.718
St. dev.	0.062	0.046	0.062	0.074	0.057	0.042	0.097	0.077	0.054	0.054	0.076	0.072	0.066
Min	0.602	0.654	0.565	0.592	0.613	0.685	0.529	0.602	0.581	0.633	0.533	0.575	0.575
Max	0.840	0.885	0.840	0.885	0.893	0.855	0.885	0.840	0.840	0.840	0.870	0.826	0.840
	Horse	Laptop	Lion	Mic	Monkey	Parrot	Piano	Rabbit	Shirt	Sunglasses	Table	Turtle	Zebra
Mean	0.708	0.682	0.767	0.690	0.671	0.693	0.680	0.719	0.699	0.668	0.719	0.706	0.717
St. dev.	0.061	0.043	0.066	0.061	0.096	0.065	0.051	0.055	0.086	0.055	0.044	0.067	0.070
Min	0.592	0.602	0.637	0.565	0.521	0.556	0.592	0.621	0.543	0.565	0.633	0.578	0.606
Max	0.833	0.800	0.893	0.813	0.847	0.800	0.806	0.893	0.870	0.813	0.800	0.826	0.906

TABLE 2 Average Bhattacharyya similarity scores (irrelevant posts).

	Bear	Cat	Chair	Dog	Dress	Elephant	Fish	Giraffe	Guitar	Hamster	Hat	Headband	Hedgehog
Mean	0.728	0.690	0.687	0.671	0.631	0.661	0.665	0.717	0.626	0.694	0.664	0.665	0.681
St. dev.	0.074	0.061	0.061	0.067	0.065	0.062	0.092	0.055	0.062	0.079	0.078	0.080	0.071
Min	0.581	0.587	0.580	0.548	0.519	0.555	0.510	0.620	0.523	0.530	0.525	0.518	0.537
Max	0.856	0.821	0.785	0.810	0.754	0.841	0.814	0.846	0.751	0.841	0.798	0.815	0.807
	Horse	Laptop	Lion	Mic	Monkey	Parrot	Piano	Rabbit	Shirt	Sunglasses	Table	Turtle	Zebra
Mean	0.706	0.619	0.673	0.620	0.691	0.622	0.707	0.681	0.596	0.686	0.688	0.713	0.694
St. dev.	0.090	0.063	0.080	0.073	0.053	0.079	0.061	0.053	0.062	0.089	0.101	0.077	0.046
Min	0.528	0.526	0.524	0.515	0.614	0.513	0.583	0.562	0.526	0.540	0.538	0.586	0.606
Max	0.889	0.797	0.825	0.820	0.814	0.828	0.824	0.799	0.758	0.809	0.858	0.836	0.795

hashtag subject #dog, we randomly select posts that depict dog(s) and posts that, despite containing the hashtag #dog, they do not contain a dog. The selection process was random, and only the confirmation regarding the visual relevance is done through human inspection.

For each pair of posts P_i and P_j belonging to the same group, we isolate their corresponding images I_i and I_j , and we computed their color histograms, expressing them as vectors \vec{c}_i and \vec{c}_j , which are then used for computing their Bhattacharyya similarity with the aid of Equation (1). Average similarity scores for both relevant and irrelevant posts were computed for each hashtag subject group of posts as shown in Tables 1, 2, respectively.

At the same time, the similarity of their corresponding hashtags, say \mathcal{H}_i and \mathcal{H}_j , was computed through the process described in Section 5.1. An inherent difficulty, however, derives from the fact that Instagram hashtags are unstructured and ungrammatical, and it is important to use linguistic preprocessing to

1. Remove stophashtags (Giannoulakis and Tsapatsoulis, 2016a), i.e., hashtags that are used to fool the search results of the Instagram platform
2. Split a composite hashtag to its consisting words (e.g., the hashtag “#spoilyourselfthisseason” should be split into four words: “spoil”, “yourself”, “this”, “season”)
3. Remove stopwords that were produced in the previous stage (e.g., the word “this” in the previous example)
4. Perform spelling checks to account for (usually intentionally) misspelled hashtags (e.g., “#headaband”, “#headabandss” should be changed to “#headband”)
5. Perform lemmatization to merge hashtags that share the same or similar meaning
6. Filter out hashtags irrelevant to visual content (Giannoulakis and Tsapatsoulis, 2019).

The aforementioned preprocessing was conducted with the help of the Natural Language ToolKit (NLTK)², Wordnet³, and in-house developed code in Python. For each token derived, its corresponding word embeddings vector representation was created using the Gensim library (see Section 5.2).

The final step was to compute the correlation between hashtag set (mean) similarities (one hashtag set for each post) and color histogram (mean) similarities with the help of Pearson correlation (Puglisi et al., 2015) coefficient for both the relevant and irrelevant posts. By rejecting the H_{01} null hypothesis, we can conclude that color similarity of images can be predicted by the similarity of their corresponding hashtag sets. Failing to reject the H_{01} , null hypothesis indicates that the information obtained from hashtag sets and color histograms, respectively, could be seen either as a complementary source for image retrieval and automatic image annotation or totally uncorrelated. This depends on the decision regarding the second (H_{02}) null hypothesis. By rejecting the H_{02} null hypothesis, we can conclude that the information provided by color histograms and hashtag sets is much more correlated in

relevant posts (as one would expect from the fact that both have the same, or similar, visual content) than in irrelevant posts.

The purpose of the study is to study the correlation between Instagram image and filtered hashtags sets. The primary purpose of the data collected (Instagram images and hashtags) is exactly to achieve the purpose of this study. Pearson correlation measures the relationship between objects. Moreover, Pearson correlation is among the most commonly used approaches (Puglisi et al., 2015). Building upon the work of Zhang et al. (2019), whom they use Pearson correlation to calculate the similarity between image and tag vectors, for rejecting the H_{02} hypothesis, it is sufficient to compare the correlation coefficient of relevant and irrelevant posts. The methods for comparing coefficient is either Zou’s confidence interval or z-score (Diedenhofen and Musch, 2015). In our case, we decided to use z-score, following the approach of other researches (Schreiber et al., 2013; Bhattacharjee et al., 2019; Younes and Reips, 2019).

5.1. Matching hashtag sets

Let \mathcal{H}_i and \mathcal{H}_j be the filtered hashtag sets of Instagram posts corresponding to the i -th and j -th Instagram images, respectively. The matching score $R(\mathcal{H}_i, \mathcal{H}_j)$ between these two sets is computed as a weighted sum of the pair similarities between the word embeddings of their constituting hashtags, as shown in Equation 6.

$$R(\mathcal{H}_i, \mathcal{H}_j) = \frac{1}{|\mathcal{H}_i| \cdot |\mathcal{H}_j|} \sum_{h_{ik} \in \mathcal{H}_i} \sum_{h_{j\epsilon} \in \mathcal{H}_j} cc(\vec{h}_{ik}, \vec{h}_{j\epsilon}), \quad (6)$$

where $|\mathcal{H}|$ denotes the cardinality of set \mathcal{H} , \vec{h}_{ik} , and $\vec{h}_{j\epsilon}$ are the word embeddings of hashtags h_{ik} and $h_{j\epsilon}$ belonging to hashtags sets \mathcal{H}_i and \mathcal{H}_j respectively, and $cc(.,.)$ is the similarity measure used with the word embeddings of Gensim models.⁴

5.2. Corpus

To evaluate the proposed methodology, along with the two null hypotheses, we randomly selected 26 independent hashtag subjects.⁵ For each one of the 26 hashtag subjects, we collected 10 relevant and 10 irrelevant Instagram posts (images and corresponding hashtags). The above process created a corpus of 520 (260 relevant and 260 non-relevant) images and 8199 hashtags (2883 for relevant images and 5316 non-relevant images). As already mentioned, relevant Instagram posts are posts whose image match the hashtag subject, and irrelevant posts are those that their image do not match the searched hashtag. Figures 2, 3 show an example of a relevant and irrelevant Instagram post, respectively, for the hashtag subject #laptop.

⁴ <https://radimrehurek.com/gensim/models/word2vec.html>

⁵ List of hashtag subjects selected: #bear, #cat, #chair, #dog, #dress, #elephant, #fish, #giraffe, #guitar, #hamster, #hat, #headband, #hedgohog, #horse, #laptop, #lion, #mic, #monkey, #parrot, #piano, #rabbit, #shirt, #sunglasses, #table, #turtle, #zebra.

² <https://www.nltk.org/>

³ <https://wordnet.princeton.edu/>



FIGURE 2 Example of a relevant Instagram post for hashtag #laptop.

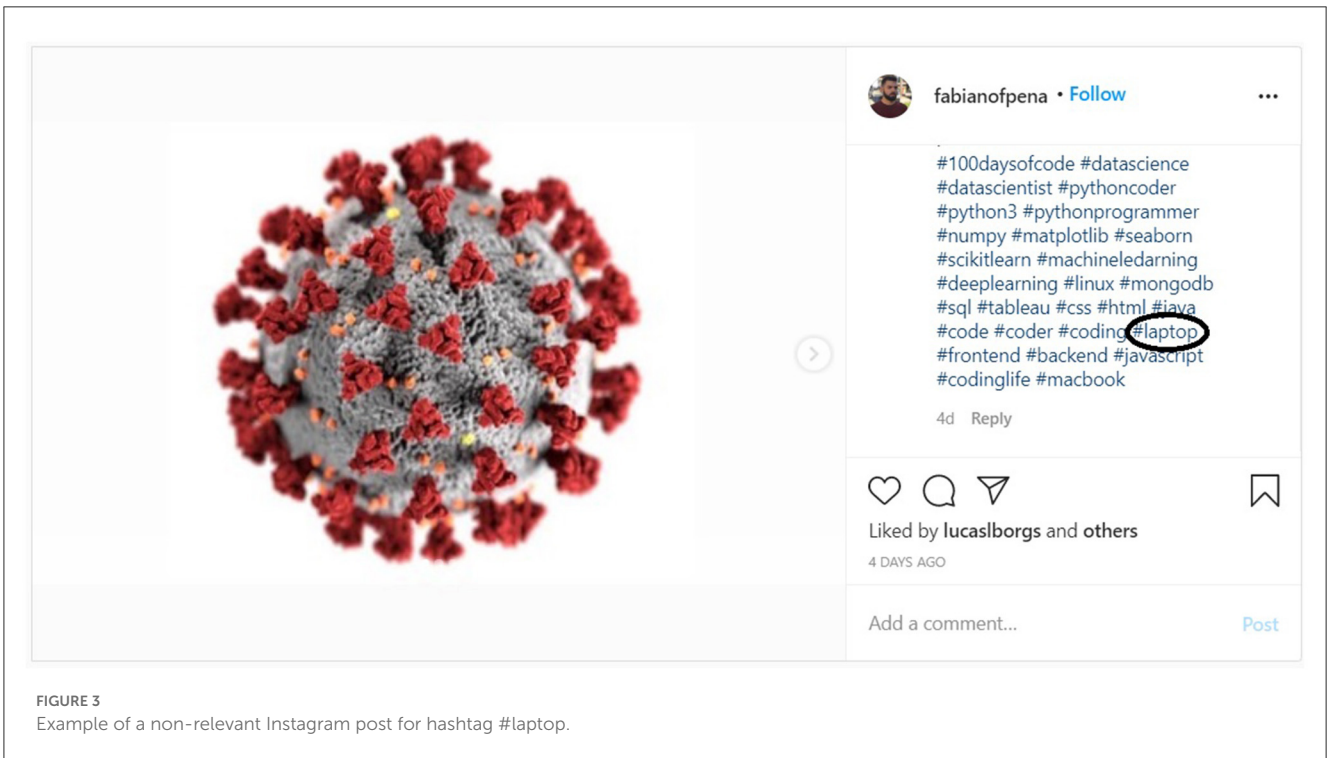


FIGURE 3 Example of a non-relevant Instagram post for hashtag #laptop.

TABLE 3 Average similarity of hashtag sets (relevant posts).

	Bear	Cat	Chair	Dog	Dress	Elephant	Fish	Giraffe	Guitar	Hamster	Hat	Headband	Hedgehog
Mean	0.350	0.255	0.263	0.224	0.250	0.247	0.167	0.197	0.254	0.257	0.294	0.189	0.160
St. dev.	0.084	0.137	0.109	0.097	0.068	0.115	0.075	0.047	0.083	0.143	0.110	0.043	0.084
Min	0.183	0.049	0.103	0.097	0.107	0.100	0.069	0.126	0.148	0.087	0.126	0.124	0.046
Max	0.575	0.781	0.560	0.471	0.404	0.643	0.356	0.316	0.534	0.920	0.633	0.297	0.457
	Horse	Laptop	Lion	Mic	Monkey	Parrot	Piano	Rabbit	Shirt	Sunglasses	Table	Turtle	Zebra
Mean	0.220	0.257	0.225	0.220	0.251	0.151	0.217	0.278	0.233	0.251	0.258	0.241	0.255
St. dev.	0.117	0.042	0.050	0.043	0.090	0.069	0.079	0.168	0.052	0.083	0.070	0.067	0.077
Min	0.059	0.162	0.137	0.114	0.044	0.043	0.090	0.080	0.140	0.138	0.109	0.133	0.130
Max	0.692	0.337	0.363	0.310	0.445	0.306	0.404	0.946	0.357	0.553	0.413	0.395	0.483

TABLE 4 Average similarity of hashtag sets (irrelevant posts).

	Bear	Cat	Chair	Dog	Dress	Elephant	Fish	Giraffe	Guitar	Hamster	Hat	Headband	Hedgehog
Mean	0.148	0.134	0.183	0.209	0.240	0.189	0.138	0.139	0.238	0.114	0.261	0.187	0.158
St. dev.	0.062	0.072	0.094	0.068	0.115	0.078	0.078	0.062	0.118	0.058	0.060	0.034	0.059
Min	0.051	0.030	0.024	0.103	0.064	0.094	0.019	0.053	0.068	0.035	0.163	0.120	0.052
Max	0.322	0.291	0.395	0.330	0.464	0.507	0.332	0.291	0.519	0.298	0.417	0.249	0.277
	Horse	Laptop	Lion	Mic	Monkey	Parrot	Piano	Rabbit	Shirt	Sunglasses	Table	Turtle	Zebra
Mean	0.094	0.161	0.210	0.187	0.121	0.102	0.178	0.181	0.164	0.104	0.178	0.143	0.119
St. dev.	0.056	0.073	0.059	0.048	0.077	0.051	0.083	0.076	0.066	0.038	0.076	0.040	0.052
Min	0.013	0.028	0.115	0.107	0.017	0.015	0.054	0.035	0.081	0.043	0.065	0.058	0.020
Max	0.206	0.300	0.361	0.303	0.274	0.257	0.388	0.352	0.367	0.178	0.336	0.220	0.241

6. Experimental results and discussion

In the previous section, we presented the methodology adopted, along with the similarities calculated for images (Table 1 for relevant posts and Table 2 for irrelevant post) and hashtags (Table 3 for relevant posts and Table 4 for irrelevant post). In this section, we utilized the aforementioned results for producing aggregates, which in turn will be used for accept or reject the two null hypotheses. In order to calculate the Bhattacharyya distance and filtered hashtag similarity, we used the Python OpenCV⁶ and Gensim libraries, respectively.

Table 5 presents the aggregated statistics (across all subjects) regarding the color histogram similarities for relevant and irrelevant posts, while Figure 4 depicts a more detailed representation per subject. Despite the fluctuations across various subjects, the aggregated mean for color histogram similarity in relevant posts is significantly higher than that of irrelevant posts ($t = 4.35$, $p < 0.01$, $df = 25$, $d = 1.22$).⁷

Similarly to the above, Table 6 shows the aggregated statistics (across all subjects) regarding the hashtag sets similarities for relevant and irrelevant posts. At the same time, Figure 5 presents a per subject comparison. Fluctuations across various subjects also appear here, as in the case of color histogram similarities. Once again, the aggregated mean for hashtag sets similarity in relevant posts is significantly higher than that of irrelevant posts ($t = 6.04$, $p < 0.01$, $df = 25$, $d = 1.71$).

Figure 6 depicts the comparison⁸ of mean similarity between filtered hashtag sets and color histograms for the relevant posts. The Pearson correlation is $r_r = 0.242$, which is lower than the critical value $r_c = 0.33$ obtained for $df = 24$, and its level of significance $\alpha = 0.05$. Thus, the H_{01} null hypothesis, that the similarity of color histograms and filtered hashtags sets in the relevant posts are significantly correlated, cannot be rejected.

Figure 7 examines the case of irrelevant posts. The Pearson correlation in this case is $r_i = -0.255$, showing that in irrelevant posts information obtained through the color histograms and the hashtags sets are contradicting. This was expected since the irrelevant datasets consisted of posts whose visual content did not match the hashtag subject.

In order to examine the second null hypothesis (H_{02}), we have to compare two correlation coefficients and check the significance of their difference, assuming a normal distribution.

The z-score of a correlation coefficient is obtained using the following formula (Yuan et al., 2013):

$$z_k = 0.5 \cdot \log\left(\frac{1 + r_k}{1 - r_k}\right). \quad (7)$$

Applying formula 7, we get $z_r = 0.247$ and $z_i = -261$. In order to test the significance of z-score difference, we need to normalize the standard error (Equation 9):

$$z = \frac{z_r - z_i}{\sigma_{z_r - z_i}} \quad (8)$$

where

$$\sigma_{z_r - z_i} = \sqrt{\frac{1}{n_r - 3} + \frac{1}{n_i - 3}} \quad (9)$$

Given that $n_r = n_i = 26$ (the number of subjects) from Equation (8) and (9), we get $z = 1.73$, which gives $p = 0.042$. Thus, the H_{02} null hypothesis is rejected at a significant level $\alpha = 0.05$.

7. Conclusion

The purpose of the study was to examine if can bridge the gap between low-level features and high-level semantic content. To achieve the aforementioned purpose, we calculated the correlation between color similarity of Instagram 26 images and their corresponding filtered hashtag sets. While no statistically significant correlation between the color histogram and the corresponding hashtag sets similarities was found, the hashtags can be seen as a complementary source for image retrieval and automatic image annotation. This is supported by the fact that the correlation difference between the similarity of color histograms and corresponding hashtag sets in relevant and irrelevant posts is both high and significant. This means that Instagram images of similar visual content, i.e., relevant posts, share similar hashtags as well. This is not the case for Instagram images of varying visual content that share few (at least one) hashtags.

Another finding of this study is that both the color histogram and the hashtag sets similarities are significantly higher in relevant posts than in irrelevant ones. Thus, it is confirmed that both color histograms and hashtag sets provide important information related to the visual content of Instagram images. Comparing the effect size using the Cohen d coefficient for color histograms and hashtag sets, one can observe that hashtag sets provide more rich information regarding the visual content of Instagram images.

Finally, we should state that we recognize that the use of color histograms might introduce some limitations to this study due to their inherent limitations (it ignores spatial relationships between pixels, it is sensitive to changes in illumination, it does not take into account the semantic meaning of the objects or scenes, and it is susceptible to noise and occlusion in images). However, we do not believe that those limitations were sufficient for changing the overall findings of this study, especially for the irrelevant posts dataset since images of the same group were irrelevant. Nevertheless, in future work, we will use different images' representations along with more images, alleviating any possible statistical bias attributed to either the images' representation or the images used. Furthermore, we plan to investigate experimentally whether the information provided by color histograms and hashtags sets is, indeed, complementary in the context of image retrieval. Our assumption is that a hybrid retrieval scheme that combines color histograms and hashtag sets will give significantly better retrieval results than each one of the individual methods alone. Finally, in an attempt to improve the correlation among the color histogram and the corresponding hashtag sets similarities, a more sophisticated filtering technique extending

⁶ <https://opencv.org/>

⁷ d is the Cohen coefficient denoting the effect size.

⁸ Equalization of means has been performed for better visualization.

TABLE 5 Aggregated statistics of Bhattacharyya similarity across all subjects for relevant and irrelevant posts.

	Mean	St. dev.	Min	Max
Relevant images	0.710	0.028	0.668	0.767
Non-relevant images	0.672	0.034	0.596	0.728

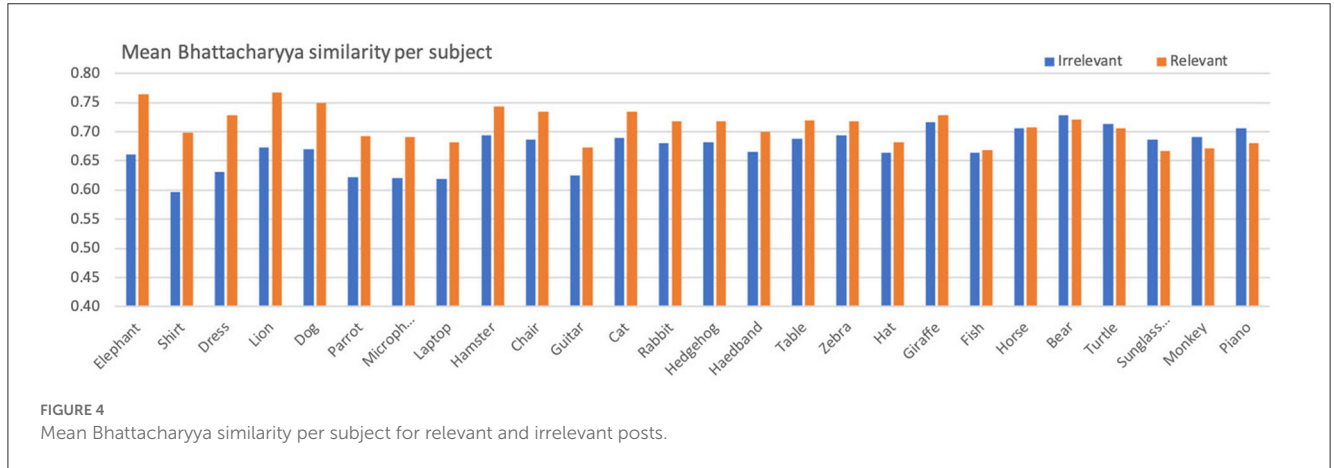
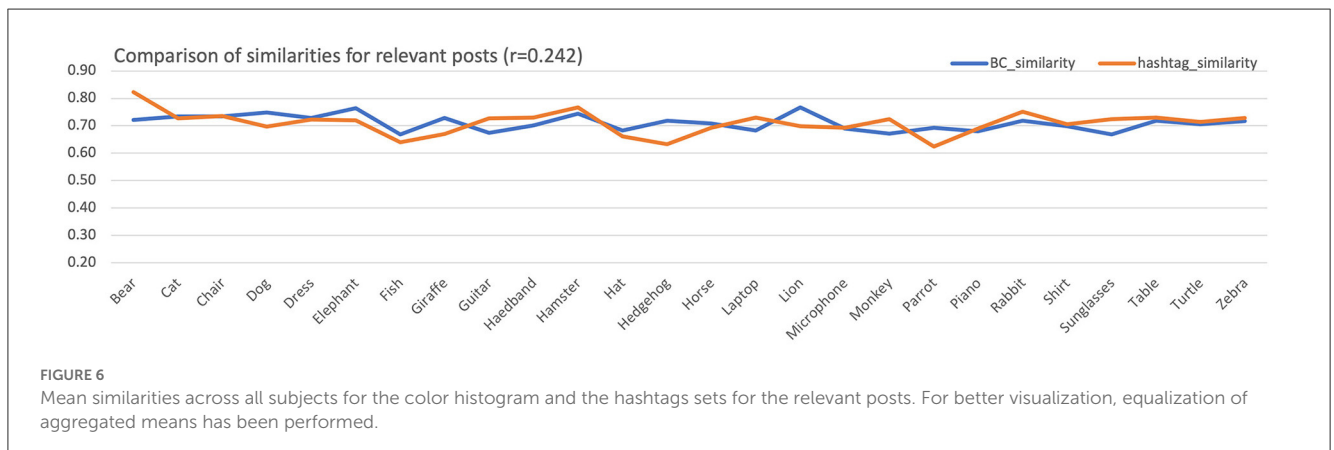
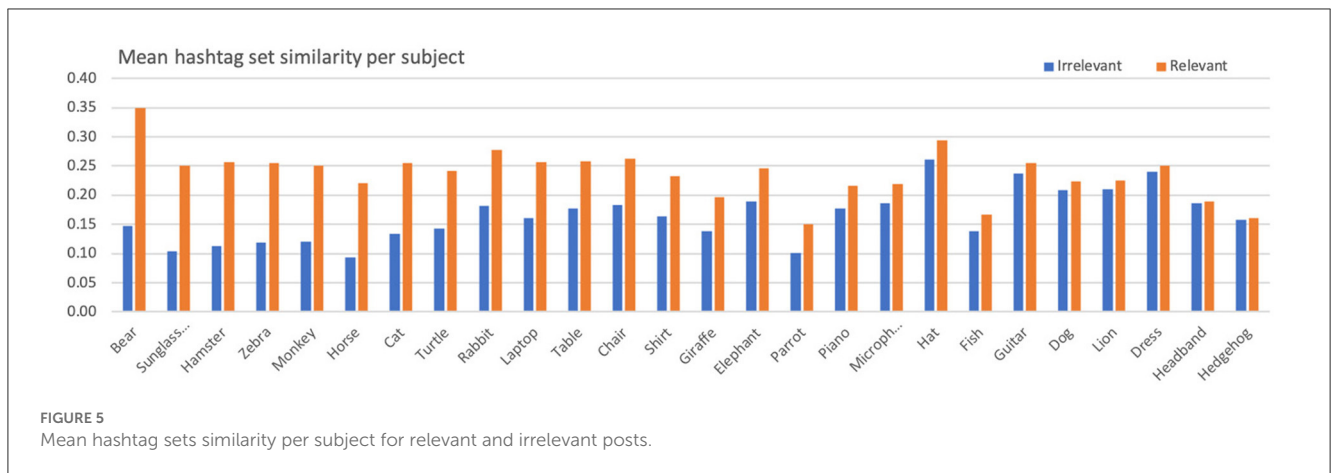
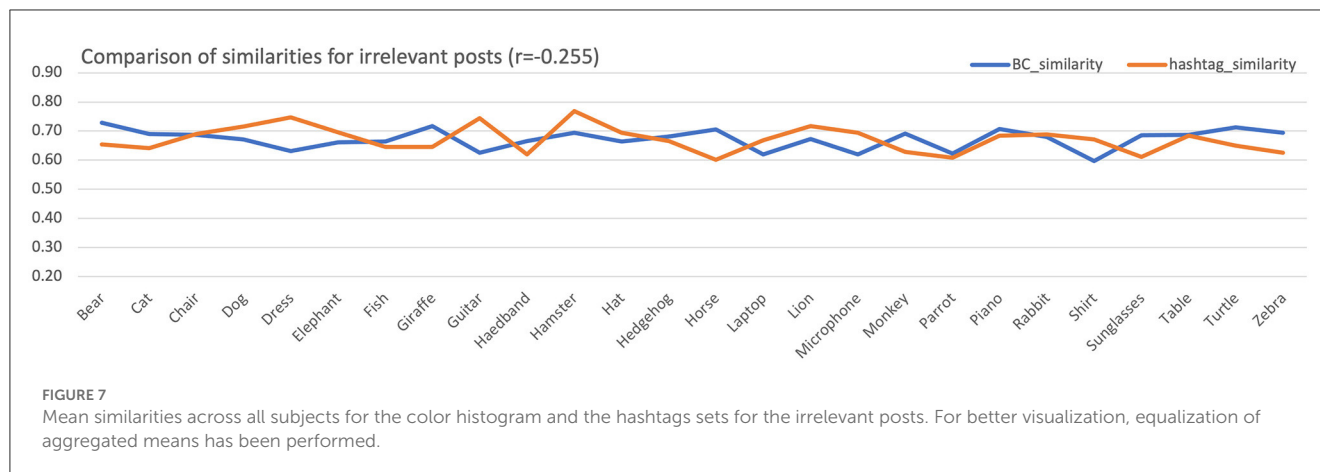


TABLE 6 Aggregated statistics of hashtag sets similarity across all subjects for relevant and irrelevant posts.

	Mean	St. dev.	Min	Max
Relevant images	0.237	0.041	0.151	0.350
Non-relevant images	0.165	0.043	0.094	0.261





our previous study [Giannoulakis and Tsapatsoulis \(2019\)](#) will be considered.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

References

- Abidi, H., Chtourou, M., Kaaniche, K., and Mekki, H. (2017). Visual servoing based on efficient histogram information. *Int. J. Control Automat. Syst.* 15, 1746–1753. doi: 10.1007/s12555-016-0070-2
- Akbar Septiandri, A., and Wibisono, O. (2017). Detecting spam comments on Indonesia's Instagram posts. *J. Phys. Conf. Ser.* 801, 1–7. doi: 10.1088/1742-6596/801/1/012069
- Arai, K. (2019). "Image retrieval method based on back-projection," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference* (Las Vegas, NV: Springer Nature), 689–698.
- Argyrou, A., Giannoulakis, S., and Tsapatsoulis, N. (2018). "Topic modelling on Instagram hashtags: an alternative way to automatic image annotation?" in *Proceedings of the 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (Zaragoza: IEEE), 61–67.
- Bhattacharjee, D., Vracar, S., Nightingale, P. G., Williams, J. A., Gkoutos, G. V., Stratton, I. M., et al. (2019). Utility of hba1c assessment in people with diabetes awaiting liver transplantation. *Diabet. Med.* 36, 1444–1452. doi: 10.1111/dme.13870
- Chacon-Quesada, R., and Siles-Canales, F. (2017). "Evaluation of different histogram distances for temporal segmentation in digital videos of football matches from TV broadcast" in *Proceedings of the 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)* (Funchal: IEEE).
- Chen, Y., Zeng, X., Chen, X., and Guo, W. (2020). A survey on automatic image annotation. *Applied Intell.* 50, 3412–3428. doi: 10.1007/s10489-020-01696-2
- Daer, A. R., Hoffman, R. F., and Goodman, S. (2014). Rhetorical functions of hashtag forms across social media applications. *Commun. Design Q. Rev.* 3, 12–16. doi: 10.1145/2721882.2721884
- Deza, E., and Deza, M. M. (2009). "Image distances," in *Encyclopedia of Distances* (Dordrecht: Springer-Verlag), 349–362.
- Diedenhofen, B., and Musch, J. (2015). cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE* 10, e0121945. doi: 10.1371/journal.pone.0121945
- Doulah, A., and Sazonov, E. (2017). "Clustering of food intake images into food and non-food categories," in *Proceedings of the 5th International Work-Conference* (Granada: Springer International), 454–463.
- Forero, M. G., Arias-Rubio, C., and Gonzalez, B. T. (2019). "Analytical comparison of histogram distance measures," in *Proceedings of the 23rd Iberoamerican Congress* (Madrid: Springer Nature), 81–90.
- Ganguly, D. (2020). Learning variable-length representation of words. *Pattern Recogn.* 102, 1–10. doi: 10.1016/j.patcog.2020.107306
- Giannoulakis, S., and Tsapatsoulis, N. (2015). "Instagram hashtags as image annotation metadata," in *Artificial Intelligence Applications and Innovations* (Bayonne, NJ: Springer), 206–220.
- Giannoulakis, S., and Tsapatsoulis, N. (2016a). "Defining and identifying stophashtags in Instagram," in *Proceedings of the 2nd INNS Conference on Big Data* (Thessaloniki: Springer International Publishing), 304–313.
- Giannoulakis, S., and Tsapatsoulis, N. (2016b). Evaluating the descriptive power of Instagram hashtags. *J. Innov. Digit. Ecosyst.* 3, 114–129. doi: 10.1016/j.jides.2016.10.001
- Giannoulakis, S., and Tsapatsoulis, N. (2019). Filtering Instagram hashtags through crowdtagging and the hits algorithm. *IEEE Trans. Comput. Soc. Syst.* 6, 592–603. doi: 10.1109/TCSS.2019.2914080

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Giannoulakis, S., Tsapatsoulis, N., and Ntalianis, K. (2017). "Identifying image tags from Instagram hashtags using the hits algorithm," in *Proceedings of the 2017 IEEE Cyber Science and Technology Congress* (Orlando, FL: IEEE), 89–94.
- Gomez, R., Gibert, J., Gomez, L., and Karatzas, D. (2020). "Location sensitive image retrieval and tagging," in *Computer Vision—ECCV 2020* (Glasgow: Springer Nature), 649–665.
- Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2018). "Learning from #barcelona Instagram data what locals and tourists post about its neighbourhoods," in *Computer Vision—ECCV 2018 Workshops* (Munich: Springer Nature), 530–544.
- Hammar, K., Jaradat, S., Dokoohaki, N., and Matskin, M. (2018). "Deep text mining of Instagram data without strong supervision," in *Proceedings of the 18th IEEE/WIC/ACM International Conference on Web Intelligence* (Weihai: IEEE), 158–165.
- Han, D. (2015). *Particle image segmentation based on bhattacharyya distance* (Master's thesis). Arizona State University, Tempe, AZ, United States.
- Jiang, K., Feng, S., Calix, R. A., and Bernard, G. R. (2020). "Assessment of word embedding techniques for identification of personal experience tweets pertaining to medication uses," in *Precision Health and Medicine: A Digital Revolution in Healthcare*, eds A. Shaban-Nejad and M. Michalowski (Hawaii, HI: Springer Nature), 45–55.
- Kayhan, N., and Fekri-Ershad, S. (2021). Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns. *Multimedia Tools Appl.* 80, 32763–32790. doi: 10.1007/s11042-021-11217-z
- Kim, Y., and Seo, J. (2020). Detection of rapidly spreading hashtags via social networks. *IEEE Access* 8, 39847–39860. doi: 10.1109/ACCESS.2020.2976126
- Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., et al. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Math. Prob. Eng.* 2019, 1–21. doi: 10.1155/2019/9658350
- Li, C., and Qian, S. (2015). Measuring image similarity based on shape context. *Int. J. Multimedia Ubiquit. Eng.* 10, 127–134. doi: 10.14257/ijmue.2015.10.3.13
- Liu, G.-H., and Yang, J.-Y. (2013). Content-based image retrieval using color difference histogram. *Pattern Recogn.* 46, 188–198. doi: 10.1016/j.patcog.2012.06.001
- Liu, K., Motta, G., You, L., and Ma, T. (2015). "A threefold similarity analysis of crowdsourcing feeds," in *2015 International Conference on Services Science ICSS 2015* (Weihai: IEEE), 93–98.
- Liu, S., and Jansson, P. (2017). *City Event Identification from Instagram Data Using Word Embedding and Topic Model Visualization*. Technical report, Arcadia University of Applied Sciences.
- Mufarroha, F. A., Anamisa, D. R., and Hapsani, A. G. (2020). Content based image retrieval using two color feature extraction. *J. Phys. Conf. Ser.* 1569, 1–6. doi: 10.1088/1742-6596/1569/3/032072
- Ong, K. M., Ong, P., Sia, C. K., and Low, E. S. (2019). Effective moving object tracking using modified flower pollination algorithm for visible image sequences under complicated background. *Appl. Soft Comput. J.* 83, 1–30. doi: 10.1016/j.asoc.2019.105625
- Pennington, J., Socher, R., and Manning, C. D. (2014). "Glove: global vectors for word representation," in *Conference on Empirical Methods in Natural Language Processing* (Doha: Association for Computational Linguistics), 1532–1543.
- Prabowo, F., and Purwarianti, A. (2017). "Instagram online shop's comment classification using statistical approach," in *Proceedings of the 2nd International conferences on Information Technology, Information Systems and Electrical Engineering* (Yogyakarta: IEEE), 282–287.
- Prieto, M. S., and Allen, A. R. (2003). A similarity metric for edge images. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1265–1273. doi: 10.1109/TPAMI.2003.1233900
- Puglisi, S., Parra-Arnau, J., Forné, J., and Rebollo-Monedero, D. (2015). On content-based recommendation and user privacy in social-tagging systems. *Comput. Standards Interfaces* 41, 17–27. doi: 10.1016/j.csi.2015.01.004
- Schreiber, R., Bellinazzi, V. R., Sposito, A. C., Mill, J. é. G., Krieger, J. E., Pereira, A. C., et al. (2013). Influence of the C242T polymorphism of the p22-phox gene (CYBA) on the interaction between urinary sodium excretion and blood pressure in an urban Brazilian population. *PLoS ONE* 8, e81054. doi: 10.1371/journal.pone.0081054
- Serafimov, D., Mirchev, M., and Mishkovski, I. (2019). "Friendship paradox and hashtag embedding in the Instagram social network," in *Proceedings of the 11th International Conference, ICT Innovations* (Ohrid: Springer Nature), 121–133.
- Sergyán, S. (2008). "Color histogram features based image classification in content-based image retrieval systems," in *Proceedings of the 6th International Symposium on Applied Machine Intelligence and Informatics* (Herlany: IEEE), 221–224.
- Takeishi, M., Oguro, D., Kikuchi, H., and Shin, J. (2018). "Histogram-based image retrieval keyed by normalized hsy histograms and its experiments on a pilot dataset," in *Proceedings of the 2018 IEEE International Conference on Consumer Electronics - Asia* (Jeju: IEEE).
- Theodosiou, Z. (2014). *Image retrieval: modelling keywords via low-level features* (Ph.D. thesis). Cyprus University of Technology, Cyprus.
- Tsapatsoulis, N. (2016). "Web image indexing using wice and a learning-free language model," in *Proceedings of the 12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2016)* (Thessaloniki: Springer International Publishing), 131–140.
- Tsapatsoulis, N. (2020). "Image retrieval via topic modelling of Instagram hashtags," in *Proceedings of the 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (Zakynthos: IEEE), 1–6.
- Tsapatsoulis, N., and Djouvas, C. (2019). Opinion mining from social media short texts: does collective intelligence beat deep learning? *Front. Robot. AI* 5, 138. doi: 10.3389/frobt.2018.00138
- Tsapatsoulis, N., Partaourides, H., Christodoulou, C., and Djouvas, C. (2022). "Quo vadis computer science? The topics of the influential papers during the period 2014–2021," in *2022 IEEE Intl Conf on Dependable, Autonomous and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (Falerna: IEEE), 1–8.
- Weston, J., Chopra, S., and Adams, K. (2014). "#tagSpace: Semantic embeddings from hashtags," in *Conference on Empirical Methods in Natural Language Processing* (Doha: Association for Computational Linguistics), 1822–1827.
- Xu, Y., Nguyen, H., and Li, Y. (2020). A semantic based approach for topic evaluation in information filtering. *IEEE Access* 8, 66977–66988. doi: 10.1109/ACCESS.2020.2985079
- Younes, N., and Reips, U.-D. (2019). Guideline for improving the reliability of google ngram studies: evidence from religious terms. *PLoS ONE* 14, e0213554. doi: 10.1371/journal.pone.0213554
- Yuan, Z., Liu, H., Zhang, X., Li, F., Zhao, J., Zhang, F., et al. (2013). From interaction to co-association—a fisher r-to-z transformation-based simple statistic for real world genome-wide association study. *PLoS ONE* 8, e70774. doi: 10.1371/journal.pone.0070774
- Zhang, D. (2006). *Statistical part-based models: theory and applications in image similarity, object detection and region labeling* (Ph.D. thesis). New York, NY: Columbia University.
- Zhang, D., Islam, M. M., and Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recogn.* 45, 346–362. doi: 10.1016/j.patcog.2011.05.013
- Zhang, H. (2019). *Dynamic word embedding for news analysis* (Master's thesis). Beijing: University of California, United States.
- Zhang, H., Jiang, M., and Kou, Q. (2020). Color image retrieval algorithm fusing color and principal curvatures information. *IEEE Access* 8, 184945–184954. doi: 10.1109/ACCESS.2020.3030056
- Zhang, Y., Wang, X., Sakai, Y., and Yamasaki, T. (2019). "Measuring similarity between brands using followers' post in social media," in *MMAsia '19: Proceedings of the ACM Multimedia Asia* (New York, NY: Association for Computing Machinery).