What you see is what you get? Automatic Image Verification for Online News Content

Sarah Elkasrawi*, Syed Saqib Bukhari
German Research Center
for Artificial Intelligence (DFKI)
Kaiserslautern, Deutschland
Email: sara_farouk.elkasrawi,
saqib.bukhari@dfki.de

Ahmed Abdelsamad*
German University in Cairo
E-11432 Cairo, Egypt
Email: ahmed.abdel-samad@student.guc.edu.eg

Andreas Dengel
University of Kaiserslautern
German Research Center for
Artificial Intelligence (DFKI)
Kaiserslautern, Deutschland
Email:andreas.dengel@dfki.de

***Abstract*—Consuming news over online media has witnessed rapid growth in recent years, especially with the increasing popularity of social media. However, the ease and speed with which users can access and share information online facilitated the dissemination of false or unverified information. One way of assessing the credibility of online news stories is by examining the attached images. These images could be fake, manipulated or not belonging to the context of the accompanying news story. Previous attempts to news verification provided the user with a set of related images for manual inspection.**

In this work, we present a semi-automatic approach to assist news-consumers in instantaneously assessing the credibility of information in hypertext news articles by means of meta-data and feature analysis of images in the articles. In the first phase, we use a hybrid approach including image and text clustering techniques for checking the authenticity of an image. In the second phase, we use a hierarchical feature analysis technique for checking the alteration in an image, where different sets of features, such as edges and SURF, are used. In contrast to recently reported manual news verification, our presented work shows a quantitative measurement on a custom dataset[1]. Results revealed an accuracy of 72.7% for checking the authenticity of attached images with a dataset of 55 articles. Finding alterations in images resulted in an accuracy of 88% for a dataset of 50 images.

***Keywords*—Media Verification, News, Journalism, Online Document Analysis**

I. Introduction

"A picture is worth a thousand words". For this reason, news reporters seek to attach relevant photos and pictures to the articles they write. The pictures they include make the article more attractive to their readers and give their text more credibility. Despite the fact that the journalists code of ethics [9] entails reporting only the truth many seek giving their articles claimed credibility by attaching manipulated media or media that belongs to another context with their articles. Sometimes this is done unintentionally such as the example of BBC attaching a photo from Iraq while reporting on massacares in Syria[2]. Another example of manipulated images presented by Alahram News, a local Egyptian newspaper, where a picture of the former president during Middle East Peace Talks was doctored to match the content of its article[3]. These are examples from reputable newspapers done by experienced journalists. Such misinformation and errors occur in much larger numbers on social media. Despite the fact that fake articles are usually followed by other articles debunking them, by the time the latter would have appeared, the false information would have gained a huge spread. Consider the example in Fig. 1 of an article published in February 2014 on a dinosaur egg that hatched in Berlin museum. The article was shared on Facebook over 59 thousand times and was included in over 1500 tweets[4]. Two weeks after the article was published, other articles emerged reporting the story as a hoax[5]. Collectively, these articles were shared around 2000 times over Facebook and included in only 20 Tweets[6], which is less than 4% of the total shares of the fake articles.

In this respect several works have addressed credibility checks on social media. Twitter has recently become an important platform for news sharing, where its most trending topics are usually concerning an important event or news story. Castillo et al. [3] consider credibility checking of Tweets in trending topics. Their approach uses supervised learning to first detect news content among the tweets. The selected news-stories are then checked for their credibility using a J48 decision trees model trained with a set of features from the tweets. To collect trending topics from twitter they used TwitterMonitor [8] and for the evaluation of their work they used Mechanical Turk with a subset of the dataset [2]. Using features from tweets such as the number of followers of the poster, links attached to the tweet or the number of shares to evaluate the credibility of tweets has been tackled by several other authors, for example [12] and [4]. Our approach differs in that it's not specific to a certain social media platform, but rather can be applied to check any online news story with an attached image to it.

Existing tools of altered image identification, such as Foto-Forensic [7] and ELA [11] identify pixel errors based on image compression. The error does not necessarily mean alteration in the image, but may be due to technical image clarity enhancement. Image clarity enhancement is permissible according to the journalists code of ethics. Warbhe & Dharaskar [14] presented a digital image forgery detection method th detects copy-move forgery when given a forged and original image.

[1]Dataset available at: http://www.dfki.uni-kl.de/~elkasrawi/Resources/datasets/article.csv

[2]http://goo.gl/kNV6OL

[3]http://goo.gl/8KzNFc

[4]http://goo.gl/QWufL6

[5]http://goo.gl/M1HVJg

[6]http://goo.gl/lmkWL9, http://goo.gl/sIdH2j

Fig. 1: An example of a fake news article being shared on Facebook.

However, alterations in online news are not limited to objects moved/copied within the image, but even if objects were added from other images, the image should be considered as altered. Image manipulations in news articles are usually spotted by finding their original version posted somewhere in other news sources. In the example of the Alahram, other newspapers reported the doctoring by posting their original article with the original image. A pre-step for verification of manipulated image, is finding other occurrences of its original version or a nearly identical version of the image online. Pasquini et al. [10] implemented a framework which retrieves news articles that match the article in question in both content as well as the attached images. Their base assumption is that the original image of the doctored one would be somewhere over the Internet. Their framework extracts both textual and visual features to detect similarities. Textual features such as title, body, date and a set of keywords and visual features in form of SURF features of the images. Their method works recursively to increase the amount of retrieved images and gives full control to use for manual inspection of the retrieved results. Our system goes a step further by providing the user with an automatic evaluation about the authenticity and alterations of images attached to news articles.
In the following sections we present our two-phase approach for automatic verification of news articles with respect to their authenticity and alterations, starting with section II. In the following section III we present our experiments results. Finally in section IV we conclude our work and present our future directions.

II. The Online News Verification System

Our news verification approach is based on analyzing the images attached to online news articles or social media stories. In the first phase, the image is checked for its authenticity and in the second for alterations from other images. For both methods, retrieving similar or nearly similar images to the image in question is necessary.
In the following, we present our approach for finding occurrences of the image online followed by a description of both verification approaches.
Finding other occurrences of the image can be modeled as a use-case of reverse image search, which is a content based retrieval technique that returns results related to a specific query image. We use this technique to discover versions of the same image, similar images and manipulated versions of the image online. We have studied several reverse image search technologies[7] and found that the Google index is said to be larger than one hundred million gigabytes with over one billion images, making the Google Image Search[8] our best option. The Google image search returns a list of matching images based on the query image. Fig. 2 shows a sample of the image search result, where for each image the result contains the url of the image, the page containing the image, the crawl date, the image resolution and the text appearing around the image. From the Google Image Search we parse the result to extract the date of appearance. Thumbnails in the results returned by Google

Fig. 2: [Google Image Search Result] Sample image returned from Google's reverse image search.

image search affect the accuracy of breaking news validation. Their crawling date is erroneous and does not match the actual date of occurrence of the image. We filter thumbnails from the result set using the image size and resolution.

A. Verification of Image Authenticity

Our assumption is, if an image appears online in conjunction with a news article and is found to be attached to another article which dates much earlier, the image might not be authentic and the news story is likely to be fake. Excluded from this are re-occurring events such as yearly concerts or conferences.
For the authenticity check, we retrieve a set of matching images to the image in question and extract their time stamps by crawling the search results. We assume the publishing date of the news article is known to the user. However, it can be retrieved by means of crawling the online article or using the different social media API. The user is then presented with the earliest appearance of the image as shown in Fig. 3.
In some cases the nature of the article requires posting of old images such as articles on upcoming concerts or posts on politicians and celebrities statements. Our assumption is, that

[7] Google Image Search, TinEye, KarmaDecay, Saucenao
[8] http://images.google.com

Fig. 3: Example output for the browser extension for image authenticity check

such events either occur periodically in the case of concerts or occur too often around a certain date in the case of speeches by famous people. To enable the user to conclude such information, we cluster the list of all occurrences of the image. This presentation enables the user to recognize if the image appears periodically or around a certain date. The occurrences are clustered using K-means clustering algorithm on their publishing date. To determine the optimal number of clusters the silhouette index (based on silhouettes [13]) is used. Hypothetically, each cluster contains similar information as related articles appear around same dates. Therefore, instead of reading about all occurrences, the user gets a general overview on how the image was used differently across time.

B. Verification of Image Alterations

Based on the assumption that doctored image have their original image somewhere over the Internet, we look for nearly similar images to the image in question and perform a hierarchical check on each image pair.

1) Image Matches: For retrieving similar images we use the "similar images" search result from Google reverse image search. Similar images are images which are visually close in shapes, colors, etc. The images are not matching by default, but require further checking to exclude images that are not versions of the queried image. The resulting images are further examined for best matches using perceptual hashing functions. The latter create hash values according to the image's visual appearance [15]. In our work, we used the Radial variance perceptual hash [5] to extract the nearly identical images from the similar image results. Hash values are generated for all similar images as well as the query image. The cross correlation between the hash values is computed to identify their similarity.

2) Edge Based Altered Image Identification: After excluding non matching images, we compare the matching images to identify possible alterations.

Scaling and Color Conversion The first step prior to comparison is converting the images to gray scale (Fig. 4), and unifying the image resolution.

Image Difference In the next step, the image is subtracted from the reference image (Fig. 5). Morphological operators are used to open the resulting image using a rectangular structure of size three pixels wide and three pixels high (Fig. 6). The morphological operation removes thin lines and single pixels

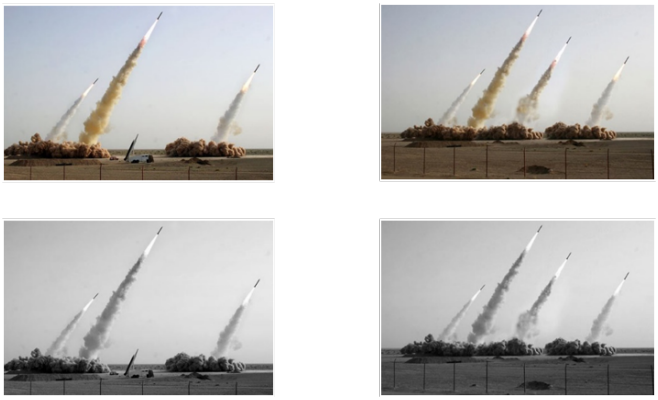

Fig. 4: Sample images showing color conversion

Fig. 5: Subtracting image after color conversion.

that may have resulted from the image scaling. The pixels in the resulting images are summed and if the sum is equal to zero; the images are considered identical.

Edge Comparison If the images are not identical, their edges are compared. We used the edge detection algorithm described by Canny which uses the first derivative operators in the process [1]. Each edge image is split into sub-images of

Fig. 6: Effect of morphological opening on the image.

thirty pixels wide and thirty pixels high. The number of edges in each sub image is summed. Then, the difference between the sum of corresponding sub images is calculated. Finally, the sum of all differences is calculated and an error e_0 is equal to total difference divided by the total amount of edges in the reference image (Fig. 7). The error value e_0 is split into three ranges:

- $e_0 > 0 \,\&\, e_0 \leq 10$ the images are considered full duplicates.
- $e_0 > 10 \,\&\, e_0 \leq 70$ the images require further checking.
- $e_0 > 80$ the images are altered.

The algorithm was tested on 30 altered image scrapped online, and the values were discovered using trial and error. Experimental results are discussed later in section III.

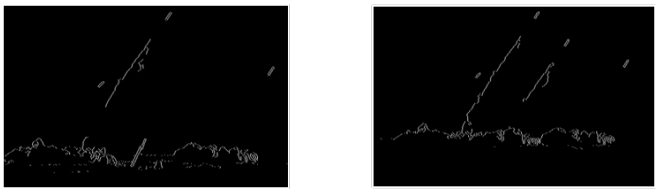

Fig. 7: Canny edge image calculation

Image Alignment If the image was not a full duplicate, we align the images by extracting SURF features from both images. A FLANN based matcher is used to match the features. The image is transformed using the feature with the minimum euclidean distance (Fig. 8).

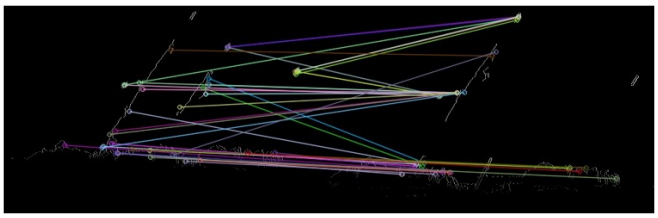

Fig. 8: Matching of SURF features displayed on the edge image

Second Edge Comparison and Decision The edges for the aligned image are calculated and compared as in II-B2. The error result e_1 of the second comparison is compared with the result of the first comparison e_0.

- if $e_1 \geq e_0$ the images are considered altered.
- if $e_1 < e_0$ the images are considered identical.

The assumption behind the decision is that if an image is identical upon alignment, the error should decrease. Therefore, if the error level increases; the image is not identical.

In order to provide the user with a friendly interface for news credibility checking, we implemented a browser extension[9] that is currently working only for Chrome browser.

[9]More info: http://www.dfki.uni-kl.de/~elkasrawi/newsverifier.html

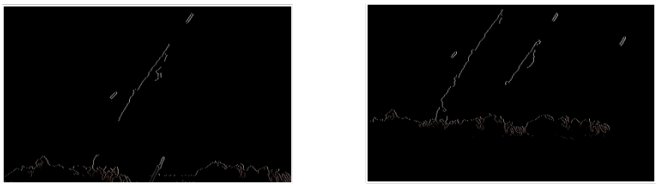

Fig. 9: [Aligned Edge Image] Canny edges are calculated for the aligned images, and altered images are identified using the new error percentage

To avoid intrusive or unwanted checks, verification is only on demand and can be performed on each image individually. The interface provides control for the user to choose which verification method is desired. A control button switches off/on the extension.

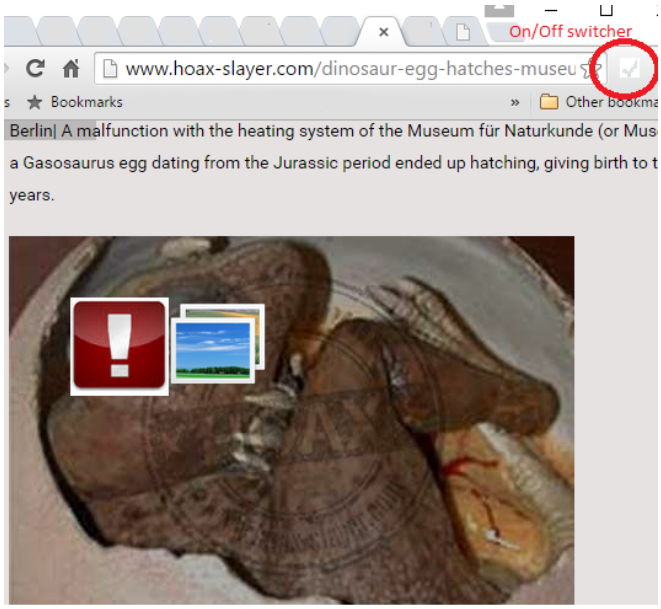

Fig. 10: Image showing extension On/Off switch as well as the two icons for verification

When the user hovers over the image in question, two control icons appear over the image as shown in Fig. 10. Based on which icon is chosen, the verification is performed. Fig. 3 shows the output after choosing image authentication option. In this example[10], a user was tweeting about the Venezuelan riots in 2013 using a picture from the Egyptian riots in 2011[11]. The system reveals the date when the image first appeared allowing the user to conclude whether the news story is fake or not. Fig. 13 shows the output after performing image doctoring check on a tweet claiming a shark tank broke in Kuwait (Fig. 11). This image is originally of escalators in a flooded underground path system in Toronto and sharks were added using photo editing software[12]. The output suggests to the user that the image might have been manipulated and a list of links to related images is presented.

[10]https://goo.gl/3F0gp4
[11]http://goo.gl/Ziyx7W
[12]http://goo.gl/JEzPxb

Fig. 11: An example of altered images in news stories

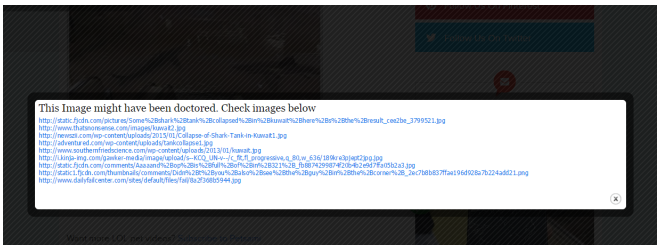

Fig. 12: Image alteration checks showing links of similar images with small modification from the image in question

III. Experiments

In this section, we discuss the methods used to evaluate both news verification phases and present the results of our experiments.

A. Dataset

Evaluating news verification on an online news dataset is a challenging task. It would require gathering a set of fake news articles or posts on social media. Most fake articles, however, are either removed after being revealed or their attached images are included in articles that debunk them which disturbs the search results [10]. We were not able to find any existing datasets of fake news. The only unbiased source of fake news was satiric websites since they clearly state, that the articles published are fake. We randomly picked articles from two websites ElKoshary[13] and TheOnion[14]in addition to a set of news stories announced as fake by the providers themselves. A set of 55 articles which contained images from these resources was collected and our algorithm for authenticity verification was tested using the implemented browser extension over the images.

Despite the fact that satiric news websites label their articles as fake, they don't clearly state if the image has been manipulated or not. For this reason we decided to rely on an existing dataset of altered images. To evaluate our algorithm for image alteration verification we used 50 pairs of images, half of which are altered the other half are identical images. Altered images were collected from the INRIA Holidays dataset [6]. Concerning identical images, we randomly selected images available online edited using the photo-editing tool Photoshop. Each image was subjected to one or more of the following edits:

- Brightness intensity
- Color maps
- Image Sharpness
- Retouching.

B. Performance of News Verification for Authenticity

The goal of the evaluation is to identify the first date of occurrence of the image and compare it with the publishing date of the article. Out of the 55 articles, 40 articles were labeled fake news because their first date of occurrence was prior to the publishing date giving an accuracy of 72.7%. The remaining 15 articles were not labeled, because the images were not indexed by the reverse image search method used.

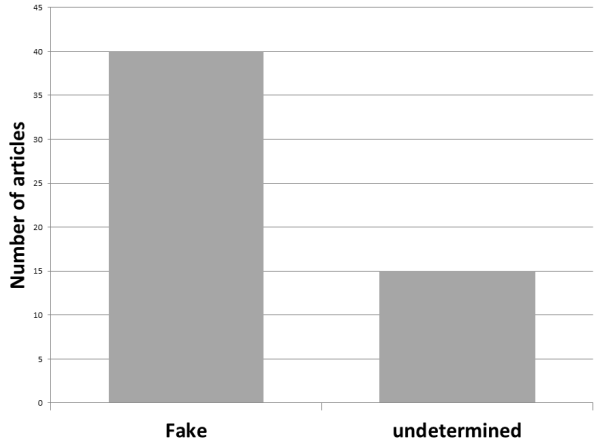

Fig. 13: Results of image authenticity verification

The difference between first date of occurrence and publishing date of the fake articles was as follows: 2 articles had a difference of more than 4 years, 10 articles had a difference between 2 and 3 years, 11 articles had a difference between 1

[13]www.elkoshary.com
[14]www.theonion.com

and 2 years, and 9 articles had a difference between 1 month and 1 year. The remaining articles had a difference between 3 days and 1 month.

C. Performance of News Verification for Alterations

In these experiments, image pairs were examined using our proposed method for alteration check. The overall accuracy to distinguish between both altered and unaltered images was 88%. The graph in Fig. 14 shows the results of the experiment. False classification of identical images was likely due to padding in the non-altered images. Upon alignment, edges are detected close to the padding area which in return increases the error value, making the image identified as altered.

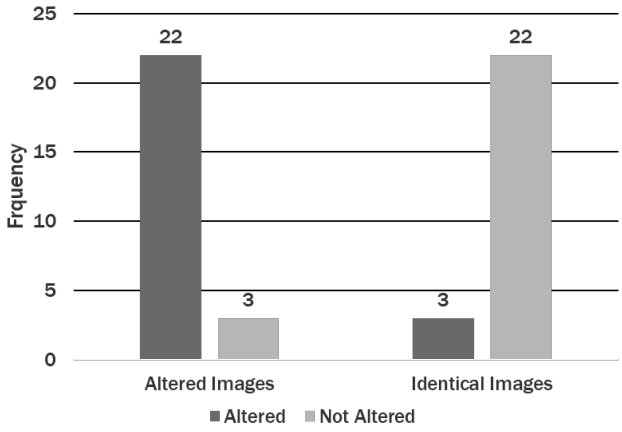

Fig. 14: Altered image identification result

IV. Conclusion

We have presented two methods for verifying online news articles and news stories over social media. Our methods make use of currently existing information online to deliver valuable content review to posts on social media.

For checking the authenticity of images attached to news stories, our algorithm resulted in an accuracy of 72.7%. Our method is a good start for assessing the authenticity of any online news piece, minimizing the damage posed by false news circulation on social media platforms. Using tools of natural language processing, further information about the news articles content can be extracted. Such information can be used to cluster the articles by their topic and not only date of appearance. Thus providing the user with additional information to verify the image.

Furthermore, we present an altered image identification algorithm which is able to search online images to find matches and identify possible alterations. Our algorithm achieved an accuracy of 88% at identifying altered images with no constraint on the type of alteration. It was also able to identify a clarity enhanced image as unaltered with an accuracy of 88%. An enhancement to the current implementation would be to highlight on the image the alterations that have been performed.

Many other features can be extracted from the images in news articles such as persons in the image, weather or some information about the location. These features can be further used for additional verification checks.

Acknowledgment

This research is supported by the Ministry for Education, Sciences, Further Education and Culture of the State of Rhineland-Palatinate (MBWWK) and is part of the project MyCustomer.

References

[1] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, Nov 1986.

[2] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. *Proceedings of the 20th international conference on World wide web - WWW '11*, page 675, 2011.

[3] M. M. Castillo, Carlos and B. Poblete. Information credibility on twitter. In *In Proceedings of the 20th international conference on World Wide Web*, pages 675–684. ACM, March 2011.

[4] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is tweeting on twitter: human, bot, or cyborg? In *Proceedings of the 26th annual computer security applications conference*, pages 21–30. ACM, 2010.

[5] C. De Roover, C. De Vleeschouwer, F. Lefebvre, and B. Macq. Robust image hashing based on radial variance of pixels. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–77. IEEE, 2005.

[6] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In A. Z. David Forsyth, Philip Torr, editor, *European Conference on Computer Vision*, volume I of *LNCS*, pages 304–317. Springer, oct 2008.

[7] D. N. Krawetz. FotoForensics:An online, real-time photo forensics system. Users can submit pictures for digital analysis and immediately see the analysis. http://fotoforensics.com/. [Online; accessed 20-December-2015].

[8] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 1155–1158. ACM, 2010.

[9] S. of Professional Journalists. SPJ Code of Ethics. http://www.spj.org/ethicscode.asp. [Online; accessed 01-April-2014].

[10] C. Pasquini, C. Brunetta, A. F. Vinci, V. Conotter, and G. Boato. Towards the verification of image integrity in online news. In *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*, pages 1–6. IEEE, 2015.

[11] pete@errorlevelanalysis.com. Image Error Level Analyser. http://www.errorlevelanalysis.com/. [Online; accessed 01-April-2014].

[12] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web*, pages 249–252. ACM, 2011.

[13] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53 – 65, 1987.

[14] A. D. Warbhe and R. V. Dharaskar. Article: Blind method for image forgery detection: A tool for digital image forensics. *IJCA Proceedings on National Conference on Innovative Paradigms in Engineering and Technology (NCIPET 2012)*, ncipet(11):37–40, March 2012. Published by Foundation of Computer Science, New York, USA.

[15] C. Zauner. Implementation and benchmarking of perceptual image hash functions, 2010.