J Braz Comput Soc (2013) 19:341–359
DOI 10.1007/s13173-013-0102-1

ORIGINAL PAPER

A survey on automatic techniques for enhancement and analysis of digital photography

Claudio S. V. C. Cavalcanti · Herman Martins Gomes · José Eustáquio Rangel De Queiroz

Received: 14 September 2012 / Accepted: 7 February 2013 / Published online: 26 March 2013
© The Brazilian Computer Society 2013

Abstract The fast growth in the consumer digital photography industry during the past decade has led to the acquisition and storage of large personal and public digital collections containing photos with different quality levels and redundancy, among other aspects. This naturally increased the difficulty in selecting or modifying those photos. Within the above context, this survey focuses on systematically reviewing the state-of-art on techniques for the enhancement and analysis of digital photos. Nevertheless, it is not within the scope of this survey to review image quality metrics for evaluating degradation due to compression, digital sensor noise, and affine issues. Assuming the photos have good quality in those aspects, this review is centered on techniques that might be useful to automate the task of selecting photos from large collections or to enhance the visual aspect of imperfect photos by using some perceptual measure.

Keywords Image enhancement · Photographic analysis · Computational aesthetics · Survey

1 Introduction

In the late 1990s, there was an immense growth in the digital photography industry. Manufacturers began to produce digital cameras on a large scale and at decreasing prices [119]. Great changes have been noticed in photographic technology and practice since then. When using consumer analog film, the number of photos was limited by the roll size (which usually allowed at most 36 photos). Nowadays, with large capacity re-writable memory cards (e.g., 256 GB), the number of photos that can be acquired/stored has increased by approximately three orders of magnitude (if considering digital images, captured with a resolution of 8 MP). Digital photography also changed the way photos were printed. With film, photos had to be developed first, in order to be seen, whereas when shooting with a digital camera, printing is no longer a requirement, once it is possible to preview images in the camera viewer or on a monitor screen, and then to decide which ones to print.

One consequence of those changes is that taking photos has become an almost costless task. Thus, the judgment of what could be a good shot and the care for adjusting camera settings for a specific scene becomes less usual for most consumers and even for some professional photographers. As a result, large amounts of photos are taken and stored daily. This causes difficulties in selecting which ones to print or to publish, e.g., in digital albums. In summary, this results in a scenario involving a large amount of stored photos from which just a small part will be printed. In this survey, consumer photos are considered the ones obtained (1) with minor adjustments in camera settings, (2) aiming at portraying daily events, and (3) barely exploring art basic techniques. On the other hand, professional photos differ from consumer ones by the use of more elaborate techniques and better equipment utilization, which might improve the photo quality. In this survey, professional photos are not necessarily obtained by a professional photographer, and do not encompass other connotations of professional photos (e.g., artistic or journalistic).

There are several recent applications for which photo processing is an essential intermediate task, e.g., photo collage [149], slide showing [44], browsing [68,85,112,150], storytelling [53], and photo summarization [28,110,117,141].

C. S. V. C. Cavalcanti (✉) · H. M. Gomes · J. E. R. De Queiroz
Universidade Federal de Campina Grande, Rua Aprigio Veloso, 882 Bodocongo, Campina Grande, PB 58429-140, Brazil
e-mail: claudio.cavalcanti@gmail.com

Samsung EX1018
Page 1

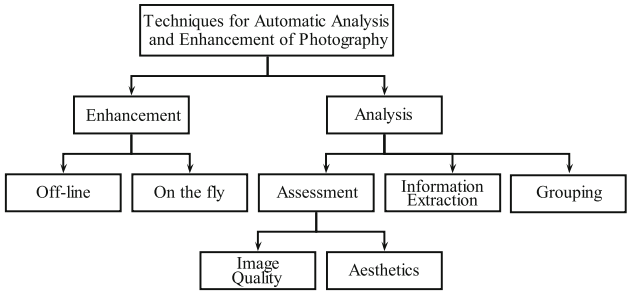

Fig. 1 Block diagram illustrating the sub-areas in which this work is subdivided

The algorithms reviewed in this survey are organized into two categories: enhancement and analysis. Each category is divided in other sub-categories. An illustration for the division that is used in this work is shown in Fig. 1.

While enhancement algorithms are intended to modify the image in such a way that it might become better-looking or appealing, analysis algorithms are designed to assess photos according to some criteria, such as composition, aesthetics, or overall quality.

A number of papers have been written on both image enhancement and analysis in the past years. This survey focuses in work on image processing techniques that were already tested or may be directly used in specific problems of photo enhancement and analysis. In order to avoid ambiguities, in this survey the words photo and photography are strictly related to consumer and professional photography.

Image enhancement algorithms can be classified as *on the fly* and *off-line*. While *on the fly* algorithms modify the photo conditions before the photo is taken, *off-line* algorithms perform changes after the acquisition took place. Although the *on the fly* algorithms might lead to better results than *off-line* algorithms, they must run faster since it might be necessary to do this in a real-time operation. *Off-line* algorithms are limited in the sense that they do not allow scene changes, e.g., it is not possible to zoom out from a photo or ask someone to open his/her eyes. However, there is no *a priori* time frame to produce the enhancement result.

Image analysis algorithms can be classified as *assessment*, *information extraction*, and grouping algorithms. Assessment algorithms analyze the visual aspect of a photo in two main facets: aesthetics and with respect to the *image quality assessment* (IQA). Formally, the main goal of IQA algorithms is to predict ratings in a human-like manner [21]. Although this definition is very broad, the term IQA is typically used to denote the evaluation of the image degradation (e.g., due to lossy compression or noise) [21,62,63,118,167,178]. Therefore, in this survey, IQA is used with this latter meaning. There is also some ambiguity regarding the use of the expression *aesthetics quality assessment*. In this survey, aesthetics quality assessment algorithms are defined as the ones whose goal is to assign a score (or a class of scores, such as professional and amateur) to a photo based in the analyzed feature, e.g., photographic composition rules or number of faces found, as used by other authors [7,74,126,145]. Moreover, information extraction algorithms search for elements of interest, such as the place a photo was taken, the existence of faces, and the presence of specific people in the environment, among others. Finally, grouping algorithms are defined in this survey as the ones which analyze images in order to find similarities between them.

It is not within the scope of this survey to review image quality assessment for evaluating degradation due to compression, digital sensor noise, and affine issues. Assuming the photos have good quality in those aspects, this review focuses on techniques that might be useful to automate the task of photo selection from large collections or to enhance the visual aspect of imperfect photos by using a perceptual measure.

This survey is organized as follows: in Sect. 2, the methodology employed for finding related work is presented. In Sect. 3, the work on image enhancement is reviewed, in particular, enhancement that could be performed to increase the quality of a photo in a printing or selection scenario. In Sect. 4, work on image (and photo) analysis are reviewed. In Sect. 5, the main issues found in the reviewed approaches are discussed and summarized. Finally, in Sect. 6, some conclusions are given.

2 Methodology of the research

This section is devoted to presenting the methodology adopted for searching the related work in the area. More specifically, information on search engines, digital libraries, and keywords used in the searching process is provided.

Two search strategies were employed, inspired by the traversing order of breadth-first and depth-first search strategies, respectively. Breadth-first search was performed within a set of predefined published conference proceedings and journals in a specified time period, but only the first level of the search tree was considered for reviewing. Depth-first search was performed by using a search engine to find papers given a set of keywords. A subsequent search was then performed by using the references of those papers as a starting point. This process was repeated until a maximum depth of 3 was reached.

In the following two subsections, more details on each type of performed search are given.

2.1 Breadth-first search

This strategy aimed at finding related papers in a set of recent technical publications, such as journals and conference

Samsung EX1018
Page 2

Table 1 Literature search results

Publication	C/J	Publisher	No. of papers
CVPR	C	IEEE	18
ICME	C	IEEE	15
CVIU	J	Elsevier	12
ICIP	C	IEEE	9
IJCV	J	Springer	8
Pattern recognition	J	Elsevier	8
MM	C	ACM	6
TIP	J	IEEE	6
IET-CV	J	IET	4
CIVR	C	ACM	3
Expert systems with applications	J	Elsevier	3
ECCV	C	Springer	2
Eurographics	C	Wiley-Blackwell	2
ICASSP	C	IEEE	2
IJPRAI	J	World Scientific	2
Transactions on consumer electronics	J	IEEE	2
Transactions on Graphics	J	ACM	2
Transactions on multimedia	J	IEEE	2
Visual communication and image representation	J	Elsevier	2
Other	J/C	-	42
Total			150

The first column corresponds to the publication name, the second column indicates if the publication is a conference or a journal, the third column shows the publisher name, and the fourth column shows the number of papers related to this survey

proceedings, given a specified time frame. The search was performed in the database of conferences and journals of IEEE, ACM, Springer, and Elsevier. Besides, by using the search results, there was also performed a search for all papers published in a given conference or journal by using its table of contents.

The keywords used for the automatic search were *(consumer OR personal OR digital) AND (image(s) OR photo(s) or photograph(s) OR photographic archive) AND (value OR quality OR aesthetics OR visual quality) AND (evaluation OR assessment OR analysis OR estimation)*. A publication period between 2006 and 2012 was defined.

2.2 Depth-first search

In this strategy, the relevant papers were found by using the following methodology: (1) based on a set of keywords, for every result returned by a given search engine, (2) the bibliography was analyzed, and (3) relevant cited work was reviewed, including the root paper itself. This is a practical and useful method for reviewing the literature, once the search is seeded using papers already considered relevant by other researchers. The great advantage is that this method dramatically reduces the searching time for finding relevant papers. On the other hand, there are some drawbacks. First, some stop criteria have to be defined, otherwise this becomes an almost endless process. Second, not every citation is directly related to the research area, since it is common to find papers from correlated areas such as Artificial Intelligence and Neurobiology.

In order to perform this search, some constraints have to be defined. The search is performed in a single level. Another level is considered if and only if a cited paper is strictly related to the area.

2.3 Search results

Table 1 contains the conferences and journals returned from the above mentioned search method. It is also noted if it is a conference or a journal, the publisher name, and the number of papers selected for this survey.

In Table 1, *Other* refers to conferences or journals with only one related publication. Figure 2 illustrates the balance between conference and journal papers that are reviewed in this survey.

Samsung EX1018
Page 3

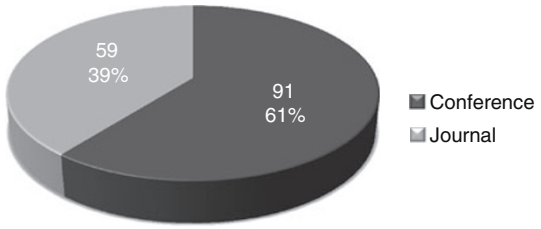

Fig. 2 Number of works published in conferences and journals that are studied in this survey.

2.4 Considerations on the methodology

The methodology previously presented was defined in order to cover relevant papers in the research taxonomy defined in the previous section. Of course, work published prior to 2006, published in low impact conferences/journals or indexed with inadequate keywords might not have been included in this survey. Nonetheless, the number of relevant papers that were included in this survey (150) indicates that a good sample of the relevant work was considered.

3 Enhancement

This section focuses on the research on enhancement techniques applied to digital photography. Usually, the areas of enhancement and analysis work side-by-side, e.g., enhancement is often performed in order to obtain more precise analysis, and a good analysis may help identify which aspects of a digital photo should be enhanced. In spite of that, and for didactic purposes, these areas are discussed separately in this survey.

As mentioned in the previous section, enhancement work may be divided in *on the fly* and *off-line*. *On the fly* approaches are the ones for which it is possible to modify the environment during image acquisition, while in *off-line* approaches that is not possible, thus, usually the photos are modified or enhanced after acquisition. Nonetheless, both expressions, *off-line* and *on the fly*, are also used with other connotations. Chartier and Renaud employed *off-line* in a noise filtering context [22], while Ercegovac and Lang used the same term in a digital arithmetic context [46].

In the following two sections, more details on photo enhancement approaches are given.

3.1 On the fly enhancement

Although it is generally possible to improve photos by means of a wide range of enhancement algorithms (e.g., red eye correction, histogram processing, among others), there are some particular scenarios from which some information is completely lost during acquisition thus making useless a post-processing operation. Photograph acquisition is naturally a lossy process, which disregards factors such as color, temperature, environment, time and space of the environment, depth of the scene, among several others. For instance, a photo may be considered inadequate due to the zoom choice, e.g., a close-up should not be used when the goal is to show that the subject is in a given location. After the photo is shot, zooming out is not possible, and a good photo might be lost. In some specific situations, it is possible to perform some correction. However, the results are usually far inferior to a scenario in which a new photo could be obtained. For example, image brightness can be adjusted after acquisition in order to improve image aesthetics, but this may result in intensity clipping.

On the fly (or dynamic, live, real-time) enhancement algorithms are proposed to automatically perform or advise adjustments to the camera settings, before the photo is taken. The performed adjustments are intended to improve the photo quality or to avoid undesired conditions, such as inadequate focus and lighting.

Most modern digital cameras have embedded *on the fly* enhancement mechanisms, such as an exposure meter, an automatic focus adjustment, and a white-balance adjustment. Since those mechanisms are mostly based on low level information, high level information about scene contents is usually input by the user by means of an appropriate *scene* switch selection. For example, one may adjust the camera *scene* switch to *motion* when using camera to shoot a sports scene.

A fully automated system may require that high-level information, such as the location of people in the scene, should be used in order to increase the overall understanding of the scene, as well as to help algorithms to decide where and how to perform changes.

A first example of a fully-automatic approach is the face-priority auto-focus [125], which has been commercially used by several camera manufacturers. The goal is to set the focus of the camera to regions where there are faces, in order to avoid incorrect focus priority. For example, the Nikon D90 camera uses face position to correctly adjust focus on people present in the scene [107].

Another example, which recently became very popular, is the smile shutter function, which shots the photo only when every detected face in the image is smiling. Several cameras and prototypes incorporating the above features have been developed by camera manufacturers, such as Sony and Canon [47].

Photographic composition rules are also considered as important features for the dynamic adjustment of an image. The conformity with photographic composition rules can be achieved with slight movements of the camera. The production of an autonomous robot photographer is a possible direction to take to address this aspect. The robot devel-

Samsung EX1018
Page 4

oped by Byers et al. [13,14] was designed to be placed in an event, moving towards possible subjects, proceeding the composition, and, finally, obtaining the photo. The subject of a photo may be identified by several approaches, such as considering the output of a skin detection algorithm and a laser range-finder sensor [13,14]. This information may also be used for finding the path that a robot must follow to reach the subjects. After getting to the desired place, once again the scene is analyzed for achieving a good composition for the photo. Four composition rules (rule of thirds, empty space, no middle, and edge rule) were used to guide their system to obtain a good composition. The system performance was assessed in some real-world events such as a wedding reception and during the SIGGRAPH 2003 conference.

Another example of a robot photographer (but less independent than the one previously presented) is the Sony Party Shot [143]. The Sony Party Shot apparatus can be plugged into the camera for locating and photographing people by moving in three degrees of freedom (pan, tilt, and zoom). The limitation of this approach is the need to put the robot in a fixed position with the aim of locating people of that point of view, and taking photos.

Some approaches may consider acquiring multiple images with different camera settings in order to detect and/or correct issues after image acquisition. For example, two subsequent photos may be obtained with different camera apertures [4]. The acquired images are then combined for finding the subject and analyzing photo composition. Besides improving photographic composition of the image, it is also possible to locate the mergers (which occur due to the projection of a 3D world scene into a 2D representation, in which background objects appear to be connected to the objects in the foreground) [4].

Despite the obvious advantage of *on the fly* approaches to digital photography, there are a number of limitations. Since most algorithms demand high processing times and some level of scene understanding, *on the fly* processing might be impractical due to battery consumption. A high processing time can be understood as a period of time that exceeds the time between two scenes setups (e.g., people position and lighting conditions).

In order to improve scene understanding, stereo vision might be employed to find interesting objects, to provide depth estimates, and to improve image segmentation as well, which may help with the acquisition of better photos. In a dynamic environment, stereo vision may be obtained by the use of Pan–Tilt–Zoom (PTZ) cameras through simple-region SSD (sum of square difference) matching [158].

Finally, *on the fly* composition might be used for automatic and semi-automatic panorama creation [23].

3.2 *Off-line* enhancement

This section is devoted to discuss existing methods for the enhancement of photos that have already been acquired. The goal is to enhance a photo using only the existing information in the image file (i.e., pixels, Exchangeable Image File format—EXIF information, faces detected, etc.). Changes typically occur at pixel level, and are required when it is not possible to obtain another photo and the resulting photo has room for improvements. For *off-line* enhancement, the image representation might be in any color space, however Benoit et al. [5] have shown some advantages when using models based on the human visual system for low-level photo enhancement.

Generally, enhancement changes can be used for making more attractive a photo that presents some type of imperfection. For instance, after removing the imperfection that lens dust may cause, a photo may look more attractive [188]. It is also possible to improve photos by smoothing the subject's skin [77], by adjusting some general aspect, such as contrast and brightness for both generic [183] and specific type of photos (e.g., improving contours in nature photos [128]) or by removing an undesired object (e.g., an unknown person or a light pole [6,170]). This last type of enhancement may be achieved by image inpainting, as discussed next.

Image inpainting has been largely used for enhancing photos [170]. Through inpainting, one might remove an element which harms the composition of a photo [170]. Inpainting algorithms remove elements from photos by evaluating the surrounding indicated area with statistical analysis and filling this area with a surrounding-like texture. Inpainting may be obtained by combining texture synthesis, geometric partial differential equations (PDEs), and coherence among neighbor pixels [6]. Patch sparsity algorithms are used for improving image inpainting [170].

Enhancement by example has also been employed [72]. When a user classifies a photo, he/she indirectly classifies the features he/she considers important. Processes such as sharpening, super-resolution, inpainting, white-balance, and deblurring are performed in photos so that they reflect the features present in example images. Other types of enhancement are the photo composite, in which a face can be replaced by another [86], and collage algorithms, in which groups selected images in a new one [149].

Cropping algorithms are designed for obtaining an image which has smaller dimensions than the original one. There are several methods for automatic [24,138,147,186,3,1,179] and semi-automatic [133] photo cropping. Cropping is performed by extracting an area of interest from an original photo [147,82,179], to improve the quality of the photographic composition [186,133,18], to retarget a photo to smaller displays [24,138,3,1,179], and to recompose the photo [87].

Samsung EX1018
Page 5

Most cropping methods have in common the use of content-aware strategies. Content detection may vary from face detection [147,24,186,138,18], saliency detection [147,24,186,3,1,87], and user-interaction with tracking of the user's gaze [133].

Region of interest (ROI) cropping methods intend to remove a fraction of the original image which contains or includes some element of interest. The dimensions of the resulting image are dependent on the original image contents. Some restrictions may apply, such as maintaining the image original proportion or leaving some room from the element of interest to the photo edges [147,82]. Common applications for such methods are thumbnail cropping and image summarization. A ROI cropping may be improved with face detection for images containing people. However, other elements of interest (e.g., animals) may be detected by using specific detectors [165] or a generic detector such as saliency maps. Saliency and spatial priors have been also used for content-aware thumbnail cropping [82].

Cropping for improving composition may help the photographer to achieve better results by modifying the image dimensions or image proportions. As it is going to be further discussed in Sect. 4.1, there are rules for analyzing the composition quality of a photo, such as the rule of thirds, which may be used in conjunction with small changes in the image dimensions, aiming at better composition by just removing a few pixel rows (or columns) of the photo. On the other hand, there are methods that make direct changes in the image contents. They will be referred to in this paper as recomposition methods, as discussed next.

Cropping algorithms are limited in the sense that they crop the images from the borders. Cropping columns or rows in the middle of the image usually results in distortions. However, in some cases the important content of the image is close to the borders. There is, however, a class of algorithms named retargeting algorithms, which may crop the image to regions other than the borders only.

Retargeting algorithms are mainly designed for adapting an image to different rendering devices, such as mobile phones [24,138,3,1]. The goal is to preserve the main content of the image while discarding unnecessary or redundant information in such a way the main content is more visible than if a simple resample were applied. Global energy optimization for the whole image may be used for image retargeting [127]. Face detection, text detection, and visual attention, all combined, may also be used for finding content in photos [24].

Another class of algorithms, similar to the retargeting algorithms is the class of recomposition algorithms. Recomposition algorithms present a very challenging area of research: the goal is to automatically change the image in order to obtain a more pleasant composition. Some examples of such changes include modifying the subject proportions, removing elements from the photo, cropping the image, etc. Most approaches are, as yet, semi-automatic in the sense they require human intervention to indicate which areas need improvement. In this survey, recomposition algorithms are considered as different from retargeting algorithms since they do not necessarily imply changing the original image dimensions or the original image proportions. The changes are usually artistic ones. Liu et al. [87] proposed a method for recomposition based on finding elements of interest, and applying composition rules (such as rule of thirds, diagonal guidance, and visual balance) to produce a better composed image. In a similar approach, Bhattacharya et al. [7] proposed a semi-automatic recomposition method which uses stress points (adapted from the rule of thirds) for optimal object placement and visual balance for improving composition. Experiments show that 73 % of recomposed images were considered better than original counterparts by human observers.

4 Analysis

In this section, relevant work on photography analysis is presented. Methods in this area may be organized according to their purposes, as follows:

1. *Assessment.* The goal is to score photos on a given scale (e.g., from zero to ten, good or bad) according to some criterion: the image quality (related to some degradation in the image) or the aesthetics of the image could be assessed; and
2. *Information extraction.* The goal is to detect the presence and location of some pre-defined elements of interest, e.g., people and faces, in a photo. The relationship between photos could also be extracted.

In the following sections, each one of these groups of methods is described in more detail.

4.1 Assessment

Assessment algorithms typically assign a score to a photo based on some metric. This allows the creation of an ordering based on the returned metric values. Assessing (or ranking) a photo is a very difficult and controversial task, especially when dealing with consumer ones. Two main aspects can be evaluated: the quality and the aesthetics of the photo. While the image quality analysis, in this survey, is understood as the assessment of the degradation of the image (e.g., sensor noise, resolution, and compression artefacts), the aesthetics analysis is related to the visual appearance and appeal of the photo (e.g., the color harmony and photo composition). IQA is out of the scope of this survey.

Samsung EX1018
Page 6

Several photo composition techniques and rules of thumb have been defined by experienced photographers, based on heuristics and are considered as responsible for improving the aesthetic quality of a photo. Those rules, known as photographic composition rules, may be used to identify higher-quality photos, based on the assessment of features. Photo composition may be regarded as the most determinant factor to consumers when considering photo quality [134].

The application of a photo composition rule will not necessarily assure best aesthetic results. Notwithstanding, photos obeying such rules are likely to look more appealing to consumers than if they were shot without attention to the rules [134,17]. However, it is not necessarily true that a photo must have composition rules obeyed to be considered appealing by consumers. This contradiction may be explained by the existence of other factors, apart from composition, e.g., people involved, photogeny and the place the photo was taken.

Some of the photo composition rules were lately explained by the theories of perception. The rule of thirds is a good example: it is known that when the subject of the photo is placed in one of the thirds of the image, the viewer is stimulated, due to the nature of human visual system, to perceive other regions of the photo. Other rules are not well defined in the specialized photography literature or are defined in terms of more subjective concepts, e.g., trying to obtain a more casual and spontaneous picture [12].

Photographic composition rules have been adopted for ranking photos in many researches [87,17,18,39,4,14,73,139,95,81,41], in which relationships between some predefined rules and the human judgment have been identified. Rules may come from human visual system theories as well as from professional photographer's expertise.

The rule of thirds is the most explored photographic composition rule in the literature [87,17,18,39,4,14,74]. One of the main reasons for that is because it is easily translated to algorithms. The rule of thirds states that one should preferentially place the subject in a third of the image width or height (depending on the image orientation). Existing works differ on how the subject of the photo is located, for example by using (1) face detection algorithms [17,18,14]; (2) low-level information, such as borders and regions found by the mean shift algorithm [87,4]; or (3) by evaluating the differences of pixels positioned in those interest areas [39].

Other rules are also explored but with less consensus between authors. The zoom rule can be applied to classify photos according to the distance from the camera to the subject. Excessive or insufficient distances are penalized by the algorithm as inadequate compositions. Since a precise detection of the subject is required for this type of analysis, Cavalcanti et al. [18] and Byers et al. [14] used face detection as the main information to identify subject position. In a similar approach, Kahn et al. [74], used the ratio between the area of the face and the area of the image.

The integrity rule was proposed to identify undesired chopping of the main subject. The great drawback of this rule is the high cost of precisely detecting the subject in a photo. The use of anthropometric measures were shown to be effective to subjects in an upright frontal position. Using some reliable information, such as the coordinates, and the dimensions of a detected face [139,18], it is possible to infer the position of the rest of the subject body, and detect possible chops.

Both zoom and integrity rules were designed considering that there is trusty high-level information such as the face coordinates. This is a great drawback, since an imprecise detection may lead to wrong conclusions. There are approaches that mainly rely on the intensities of pixels rather than on high-level information. One disadvantage is that color images may have their channels treated independently and may result in redundancy which must be treated by classification algorithms. For instance, in the work of Datta et al. [39], 56 initial rules were reduced to just 15 rules after the execution of support vector machines (SVM) and pruning [33], since there was redundancy in applying the same data extraction algorithm in all images.

Finally, the visual balance rule is also used for analyzing if the photo elements are well balanced, i.e., are placed in a photo in a way that observer attention is equally divided by the photo elements [87,7,176].

Besides the use of photographic rules there are authors which evaluate low-level features (e.g., sharpness, brightness, and contrast) in order to identify the overall appearance of a photo [108,73,39,81,80].

Higher level image analysis may also be employed for photo ranking. For instance, aesthetic analysis, may be achieved by learning how humans classify photos according to some subjective criteria. Although that might be difficult, there are studies focusing on the emotions evoked by artwork in humans [174]. The criteria may be diverse. For instance, the time a human spends evaluating an image can be a criterion for confidence on human assessment [45]. It is believed that the emotions evoked by a natural image can be understood by means of aesthetics gap concept. According to Datta et al. [40], "The aesthetics gap is the lack of coincidence between the information that one can extract from low-level visual data and the interpretation of emotions that the visual data may arouse in a particular user in a given scenario.". Color harmony can also be considered as an important feature to be considered [176]. Low-level information such as lighting, color [95,80,74], luminance [74], edges, and range of lightness [48] are used for judging the harmony (a high-level subjective aspect) of a photo and videos [104].

Samsung EX1018
Page 7

Besides all the above presented factors, there are some other common sense factors that might influence human judgement. Below is a list of these additional factors:

- *People involved.* A photo may be considered more or less appealing depending on the identity of the shown people, e.g., even a badly composed and illuminated photo might be considered good if it contains people for which the consumer has affection, such as the photographer's child, a famous person, etc. The opposite might also happen: a well composed photo might be discarded if the person in the photo is unknown;
- *Place where the photo was taken.* Some photos are related to places rarely visited. Thus, even if a photo has problems, e.g., in composition or illumination, it is likely that it will not be discarded because of its uniqueness;
- *Photogeny.* Well-composed photos do not necessarily contain photogenic people. It is possible to find one or more group members talking or looking elsewhere in the moment the photo was shot, especially in group photos; and
- *Personal preferences.* Some people might prefer a photo without obeying composition rules.

Despite the above discussed factors, photo ranking might be useful for helping consumers to identify (at least in a group of pre-selected photos) the ones with more attributes related to a better looking or appealing impression.

4.2 Information extraction

This section includes a discussion on approaches for extracting elements of interest that might be important to a photographic analysis system. The reviewed work involves approaches for face and people detection, landscape analysis (e.g., horizon tilt evaluation), and identification of the image class (e.g., if it is a photo or a graphic image). The goal is neither to rank nor to classify the images but to extract information. This may be considered as an auxiliary source of information for image ranking methods (discussed in the previous section). Elements of interest might be anything the user is searching for: (1) a face; (2) a person; (3) regions with unwanted features such as dissection lines [139], (4) unfocused or blurred regions [151], (5) a sunset area [9], (6) text [153], and many others.

Generally, information extraction by different approaches involves the construction of a classification model for the targeted element (this can be performed, for example, through a learning process using a set of reference patterns). It is commonly accepted that the best technique to build a particular model for a given problem is dependant on specific features of the problem.

Decision trees [130,43] are typically employed to identify classes that have a reduced number of constraints, both numerical and categorical, such as number of colors, number of people, etc. The ID3 classification algorithm [124] was used for classifying an image as either a digital photo or as artwork (e.g., a logo, a drawing, and other images artificially generated). The decision tree was trained with 1,200 images. An accuracy of 95.6 % was achieved when distinguishing the classes. This result was verified through a tenfold cross validation.

The SVM [33] is largely used for classification, which is useful for detecting features in photographs, such as indoor or outdoor scenes [137], the presence of a sunset [9], the level of expertise of the photographer [152], the presence of skin regions [64], among others. The SVM is normally used when the set of constraints is not small, and there is no clear linear separation of the data for each class. When defining an SVM model, a kernel must be specified. For instance, Serrano et al. [137] used a radial basis function. On the other hand, Boutell et al. [9] and Li et al. [80] used a gaussian function.

When there is a great amount of data, and a great number of components as well, the high correlation between those components may harm classification. Many authors use principal component analysis (PCA) [65] for reducing the dimensionality of the feature space [152].

Some information extraction methods are designed for detecting human-related information, such as face, eye, skin, pose, etc. Since most photos have people, such information is very important to any photographic analysis system.

Recent work in face detection focused on multi-view, rotation, and scale invariant face and eye detectors. Discriminant features [162], low-level features [42], Sobel edge detection, morphological operations, and thresholding [154] may be used for this goal.

Face recognition algorithms can be used for identifying photos which contain or do not contain a specific individual, as well as for finding relationships between images due to the presence of a given person or group [55]. The identification of a specific person might be used, according to a rule defined by Loui et al. [92], to infer the relevance of a given photo based on the relationship of the people to the photo owner. Recognition can be also used, along with human tagging and some logic formalism (e.g., Markov logic), to retrieve social connections in photo repositories [187,75,140,57].

In a photo selection scenario, it is very important to identify the relationship between the people present in some image set. Such a relationship may be used for predicting the significance of such images to that set [91]. This can be done using local patterns of Gabor magnitude and phase [166]. Face recognition highly relies on face detection. Thus, imprecision on face detection may result in a poor face recognition. There are, however, approaches for misalignment-robust face recognition [171].

Samsung EX1018
Page 8

The use of face details, such as birthmarks [120] and clothes [54], are also used to improve face recognition. A Markov random field is used for recognizing people based on contextual clues such as clothing [2]. Gender can also be a clue for face recognition by means of spatial Gaussian mixture models (SGMM) [84].

Considering that low-level image features are considered by consumers to determine if a given photo is better than another [137], detecting the presence of such low-level features may be very useful for ranking photos. One of those features is the blur. Blur may be used for automatically ranking photos [73,95]. Blurred images can be identified by the detection of some features, such as image color, gradient, and spectrum information [88]. The spectral analysis of an image gradient is also used for identifying blurring kernels in images [69]. Other features such as clarity, complexity, and color composition are also explored [94,39].

Besides the face, skin regions are another important evidence for the presence of a human in a photo. Several approaches were proposed. Skin tone may be detected by a pixel-wise approach or by a region-based approach [76]. Both approaches use a color model [64]. Additionally, it might be possible to decompose skin tone in hemoglobin and melanin [169], which can be used for a better understanding of skin texture.

Skin classification may be performed by the use of SVM and region segmentation [64], as indicated earlier in this section. The approaches might be compared with receiver operating characteristic (ROC) analysis [136].

While evidence for people can be obtained by skin detection or face detection, there are also approaches by which humans can be directly detected in images. Recent approaches use local binary patterns (LBP) for human detection through two variants, semantic-LBP and Fourier-LBP [105]. People detection can also be achieved by the use of quantified fuzzy temporal rules for representing knowledge of human spatial data. This kind of data is learned with an evolutionary approach [106]. A head and shoulders detector can be also achieved by the use of a watershed and border detector, whose outputs are used to train a classifier using AdaBoost [168].

Some researches have been conducted for detecting people in a specific context, but might be extended to a more general scenario. For instance, it was shown that the use of region covariance features with radial basis function kernel SVM, and histograms of oriented gradients (HOG) with quadratic kernel SVM outperformed the use of local receptive fields with quadratic kernel SVM in the specific scenario of pedestrian detection [116]. In the same way, the detection of human activities, such as 'fighting' or 'assault', are recognized and encoded by using context-free grammars through a method which uses a description-based approach [131].

Besides people detection, other types of information may be useful for photography analysis. For example, the social context might be inferred by analyzing the distribution of people found within the image [56]. A graph-based approach has shown to be useful for finding rows of people [56].

The pose of the people is also important information about the photo. Each body part has a limited number of positions when compared relatively to other body parts. For instance, the head is directly connected to the shoulders and might not appear connected to the feet. Thus, if a face is found, the shoulders should come right below. There are several approaches for human pose estimation. Human pose may be estimated in video sequences using multi-dimensional boosting regression from Haar features [8]. In static images, pose can be classified through angular constraints and variations of body joints with the use of SVM [97], with observation a driven Gaussian process latent variable model (ODG-PLVM) [61] and non-tree graph models [70], and with a conditional random field (CRF) if multiple views are available. A bottom-up parsing approach can be used to recognize the human body for performing pose estimation by segmenting multiple images.

Besides human subjects, other types of subjects may be considered in a photo. Different subjects (e.g., natural or man-made objects and animals) might appear alone or interacting with humans, resulting in a more complex photo. A shape-driven object detector [129,59], SIFT [78], and sets of mattes [144] may also be employed for a more general object detector. It is also possible to identify the region-of-interest by using captured camera information stored in EXIF [83].

Instead of detecting a specific type of object, it is also effective to use the identification of regions within the image with some correspondence. In this sense, the image segmentation algorithm has fundamental importance for the photography analysis.

There are several approaches to image segmentation. Since in photography analysis, subjects and scenarios might vary widely, the more general the image segmentation algorithm is, the better the result.

Main methods for image segmentation are based on edge information [159,19], fragment-based approaches [36,79], point-wise repetition [182], tree partitioning under a normalized cut criterion [160], a nonparametric Bayesian model [115], a geometric active contour model [180], Markov random fields with region growing [123], Markov random fields and graph cut [25], and local Chan–Vese (LCV) model [163].

Most algorithms deal with both color and gray images. Some image segmentation algorithms are specific to color images [19,181]. It is normally difficult to compare different image segmentation algorithms, but unsupervised objective assessment methods have been attempted for this task [185].

Samsung EX1018
Page 9

4.3 Grouping

Photo grouping is designed for setting associations between groups of photos. The associations may be set by either the semantic information found (such as number of faces detected, number of colors, etc.) or high-level information (e.g., Global Positioning System, GPS, position present in some image EXIF).

4.3.1 Classification and clustering

Classification algorithms are designed to identify the class which a given image belongs to. There are several goals for image classification, e.g., (1) identifying, in a set of image files, which ones are photos and which ones are graphics [114]; (2) identifying whether photos were obtained in an indoor or an outdoor environment [137]; and (3) identifying if images were obtained by an amateur photographer or a professional one [151,80,94,111], among others.

It is not completely known how humans perform classification tasks. Vogel et al. [157] have shown, however, that humans use both local and global region-based configurations for scene categorization. This implies that human-inspired algorithms may consider both local and global region-based information for better results in image classification. It was also shown that the color plays an important role in image categorization for humans. Natural images were better classified when presented in color as opposed to gray levels [157]. Classification has a close relation to information extraction, as discussed in Sect. 4.2.

One of the main steps for an accurate image classification is the representation of the image which will later be used as input to a classifier. Representation can be performed by local descriptors [101], a topic histogram using probabilistic latent semantic analysis (pLSA) and expectation–maximization (EM) [93], multilevel representation [164], triangular representation, which is robust to viewpoint changes [67], resolution invariant image representation [161], and scale invariant feature transform (SIFT) [164], among other methods.

Some relevant classifiers proposed in the literature are: AdaBoost [93], SVM [101,164], multiple kernel learning (MKL) [93], Bayesian belief networks [38], Bayesian active learning [122], and conditional random fields models [15,184].

The main challenges in image classification are the computational cost and the classification accuracy. A local adaptive active learning (LA-AL) method was used for lowering the number of training samples needed [93]. The within-category confusion can be dealt with probabilistic patch descriptors, which encodes the appearance of an image fragment, and the variability within a category [101].

Clustering algorithms are intended to automatically group images when considering their extracted features. Given a set of photos, clustering can be used to identify the existing relationship between such photos.

Cooper et al. [30] presented an automatic temporal similarity-based method using EXIF data. Graph-based algorithms [52] and local discriminant models and global integration (LDMGI) [173] are common methods for image clustering.

Since clustering is not commonly a supervised process, system improvements are necessary for reducing errors in the system. Thus, user feedback is used as a way to bring out relevant feedback about system performance [11].

4.3.2 Summarization

Another recent area of interest is finding relationships between photos for producing summaries. Summaries are useful since finding information in large sets of images can be time consuming. Summaries are used for producing condensed displays of touristic destinations [117], simplifying photo browsing on personal collections [141,155,28], indexing [156], and storytelling [53,110], among other applications. A specific problem related to the task of producing summaries or filtering out redundant information from a collection of photographs is the detection of near duplicates [28,148,126].

Photos matching specific keywords [142] or GPS-tagged information [135] have been grouped to build 3D models of some sightseeing. On-line tools, such as Bing maps [102], used some of those technologies for building 3D models of such places.

4.3.3 Image retrieval

According to Marshall et al. [99], image retrieval techniques can be classed as content-based image retrieval (CBIR) and annotation-based image retrieval (ABIR). In CBIR, the images are processed for obtaining information while in ABIR, images are often annotated with textual information, such as place, time or photographer, and this information is used to retrieve images.

Most detection algorithms can be used as an intermediate step for retrieving images in CBIR [99,90], such as recognized faces [187,121] and events [37], among others.

Since manually tagging photos can be time consuming, recent work considers the use of information automatically obtained from EXIF [83,148,85,132], SIFT [148,26], face recognition and connections found in social networks [27,113,155], and georeferences, which might be obtained from GPS devices [16], people clues such as faces and clothes [146] or other high-level information [132].

Samsung EX1018
Page 10

4.4 Discussion

In this section, algorithms for photo analysis have been organized in three categories: assessment, information extraction, and grouping. From the performed review, assessment seems to be the less explored area. This may be explained by the highly subjective nature of the task, which makes it difficult to perform precise or universal analyses. The other two areas are more explored in the literature and present a richer set of approaches.

Besides the underlying limitations discussed in the next section, the approaches seem very promising to be included in a photography analysis system.

5 Critical analysis

The main issues covered in the studies reviewed in this survey are considered in this section. To better discuss such issues, the following information about the articles were summarized: the source of the used image set, the size of the image set, the main goal, the metrics used for assessment, and the achieved results. The photo analysis algorithms are shown in Table 2 and the enhancement algorithms are in Table 3.

This section contains two subsections. In the first one, a review of the image sets used in the experiments is given. In the second one, commentaries about the validation processes are presented.

5.1 Image sets

For most of the image analysis algorithms reviewed in this survey, the purpose is to perform tasks in a human-like manner. Thus, it is fundamentally important to ensure the photo sample is representative for testing.

Some studies were performed to identify the user behaviour when photographing [96], sharing [103], analyzing [49], and managing [35]. However, based on the conducted literature review, strong evidence about the user preferences were not drawn, the assessment of most algorithms for photo enhancement and analysis are performed by means of subjective assessment.

According to the conducted literature review, there is no defined methodology for carrying out subjective experiments for photos analysis. Some methodology ought to be employed due to the number of factors that might influence subjective assessment. Some of these factors are:

1. *People involved.* While in professional photos, the people present in the photo are usually part of the subject, in consumer photos, people are mostly known and significant to the photo owner. Therefore, a photo assessment performed by consumers might be too strict in the absence of a known person and too flexible in the occurrence of, for instance, a family member;
2. *Place and event.* In some situations, the photo might not be technically good, but captured a place or a rare event. This could positively influence the judgement of the photos;
3. *Style used.* Different users adopt different photo habits. The individual style of a user might not be appreciated by other users;
4. *Number of Images.* There are an endless number of poses, camera settings, and subject positioning. Therefore, it is barely infeasible to represent this diversity of possibilities in a small set of images.

In Tables 2 and 3, the second column (Image sources) indicates the databases from which the images were obtained. In this column, *Web* refers to web crawled images and *Own* refers to particular photos from the authors or contributors. The third column, (Set size), represents the number of images used in the experiments (if any). The fourth column indicates the main goal of the work. The fifth column briefly indicates how the approach has been evaluated, in which *Obj.* represents an objective assessment and *Sub.* a subjective assessment method. In the final column is shown the best reported performance of proposed algorithms. Tables 2 and 3 have been built based strictly on what was described in the papers. Whenever the information was not explicitly shown in the paper, results are shown in a non-numeric way or the information is not suitable for the discussed problem, *NI* (Not Informed) is used. Both tables are sorted based on the total number of images and then alphabetically by the name of authors.

By analyzing the Tables 2 and 3 it is possible to draw some conclusions about the number of images and their sources. First, there is no consensus on the database to be used. This makes it impossible to perform a direct comparison between the results in Tables 2 and 3, and to reproduce the experiments as well. Second, the number of images employed in the evaluations drastically vary. The average number of images employed in photo analysis work is 45,344 with a standard deviation of 212,556, and a median of 3,581 with an interquartile range of 12,278. If only photo enhancement work is considered, the average is 716 with a standard deviation of 952, and a median of 375 with an interquartile range of 766. Third, no work presented a categorization of the image set, e.g., not known is the distribution of the number of people among a given set. Finally, some papers only presented a simple visual verification of the results (e.g., Achanta et al. [1] and Banerjee et al. [4]).

It is important to highlight the non-utilization of a labeled and representative public image database for photographic

Samsung EX1018
Page 11

Table 2 Summary of the reviewed work on analysis techniques

Authors	Image sources	Set size	Main goal	Assess. method: used metrics	Results
Liu et al. [90]	Photosig [10], NUS-WIDE [29], Kodak [172]	1,300,000	CBIR	Obj.: precision	14.5 %
Li et al. [83]	NI	70,000	ROI Detection	Obj.: precision and Recall	NI
Pang et al. [117]	Flickr [50]	50,000	Grouping	Sub.: scaled (1–5)	Average rank > 4
Sinha [141]	Flickr [50], Picasa [58]	40,000	Grouping	Obj.: JS Divergence	JS Div. < 0.3
Tong et al. [152]	Corel [32], MS	29,540	Assess.: home user x photographer	Obj.: MSE	11.1
Marshall [99]	MIR FLICKR 25000 [51]	25,000	CBIR	NI	NI
O'Hare [113]	Own	23,774	Grouping	Obj.: H-hit rule	NI
Dao et al. [37]	Picasa [58]	19,101	Grouping	Obj.: F-Measure	NI
Luo et al.[94]	Web	17,613	Assess.: high x low quality	Obj.: accuracy	95 %
Yao et al. [175]	Photo.net [60]	13,302	Assess.: ranking	Sub.: scale (0–100)	75.33 %
Ke et al. [73]	DpChallenge [20]	12,000	Assess.	Sub.: scale (1–10)	72 %
Yeh et al. [177]	DpChallenge [20], Flickr [50]	12,000	Assess.: ranking	Sub.: scale (1–10)	81 %
Yeh et al. [176]	DpChallenge [20], Flickr [50]	12,000	Assess.: ranking	Sub.: scale (1–10)	93 %
Sandnes [132]	Own	7,672	Grouping	Obj.: accuracy	88.1 %
Su et al. [145]	DpChallenge [20]	6,000	Assess.	Sub.: scale (1–10)	92.06 %
Boutell et al. [9]	Corel [32]/Own	5,770	Class.: sunset	accuracy	96.4 %
Singla et al. [140]	Own	4,500	Summ.	Obj.: precision and Recall	NI
Oliveira et al. [114]	Web	3,700	Class.: photo x graphic	Obj.: cross-validation	95.6 %
Datta et al. [39,41]	Photo.net [60]	3,581	Assess.: ranking	Sub.: scale (1–7)	70.12 %
Obrador et al. [111]	Photo.net [60]	3,141	Class.: high x low aesthetics	Sub.: scale (1–7)	66.5 %
Zhang et al. [187]	Own	2,597	Grouping	NI	NI
Tong et al. [151]	Corel [32]	2,355	Class.: blur	Obj.: Accuracy	98.6 %
Obrador [108]	NI	2,000	Assess.: ranking	Sub.: 6 grades	37.5 %
Shen et al. [139]	Web, Flickr [50]	2,000	Detect.: dissection lines	Sub.: TP + FP	80.87 and 33.61 %
Cooper et al. [30]	Own	1,449	Class.: event	Obj.: F-Measure	0.8568
Serrano et al. [137]	Web	1,200	Class.: indoor x Outdoor	Obj.: accuracy	90.2 %
Chu et al. [26]	Own	1,199	Grouping	Obj.: precision	0.68
Chu et al. [27,28]	Flickr [50]	1,024	Grouping	Sub.: scale (1–5)	Satisfaction > 4
Tang et al. [148]	Picasa [58]	975	Grouping	Obj.: precision and Recall	NI
Loui et al. [92]	NI	943	Grouping	Sub.: correlation	0.84
Lo Presti et al. [121]	Gallagher [54]	589	Retrieval	Obj.: error rate	27.68 %
Kim et al. [75]	Own	564	Grouping	Obj.: Precision at Top-N	MAP > 0.4
Li et al. [81]	Flickr [50]	500	Assess.: ranking	Sub.: choice	51 %
Li et al. [80]	Flickr [50]	500	Assess. & Class	Sub.: scale (0–10)	Residual sum-of-squares: 2.38
Khan et al. [74]	Li et al. [81]	500	Assess.: ranking	Sub.: choice	61.10 %
Jiang et al. [71]	Flickr [50], Kodak [172], Own	450	Assess.: ranking	Sub.: scale 0–100	MSE < 17
Obrador et al. [110]	Own	200	Grouping	Sub.: choice	75 %

Samsung EX1018
Page 12

Table 3 Summary of the reviewed work on enhancement techniques.

Authors	Image sources	Set size	Main goal	Assess. method: used metrics	Results
Byers [13]	Own	3,008	In-camera photo composition	Sub.: user selection	35 %
Tian et al. [149]	Own	1,627	Photo Collage	Sub.: professional	Most results considered good
Liu et al. [87]	Web	900	Recomposition	Sub.: forced choice	93.7 %
Bhattacharya et al. [7]	Web	632	Recomposition	Sub.: forced choice	93.7 %
Yin et al. [179]	Own	600	Media Adaptation	NI	NI
Suh et al. [147]	Corbis [31]	150	Cropping	Sub.: recognition time	Faster using the approach
Zhang et al. [186]	Own	100	Cropping	Sub.: scaled	41 %
Chen et al. [24]	Web	56	Recomposition	Sub.: scaled	71.28 %
Santella et al. [133]	NI	50	Cropping	Sub.: forced choice	58.4 %
Setlur et al. [138]	NI	40	Retargeting	Sub.: forced choice	89.1 %
Achanta et al. [1]	Berkeley [100] and MSRA [89]	NI	Retargeting	NI	NI
Banerjee et al. [4]	NI	NI	Recomposition	NI	NI
Lim et al. [86]	NI	NI	Composite	NI	NI

analysis. Therefore, most authors crawled images from online repositories. Web crawlers can be employed for creating image sets which present a richer and diverse number of situations, and a higher number of pixels [114,137,24,139]. Nevertheless, the great drawback is the lack of copyright licenses for public experiments. There are some public image databases that are free for academic research use (such as Flickr [50] and other databases under Creative Commons license [34]), yet they are not labeled. Regarding photo analysis, there are some web databases which have been used as a ground-truth for subjective quality analysis (e.g., DPChallenge [20] and Photo.net [60]). However, since those databases were designed for photo contests, they typically do not represent the reality for consumer photography, which usually have less quality and less exigent evaluators.

Two authors have built datasets in order to make them available to the community. The first work, from Luo et al. [94], presented a dataset of 17,000 labeled photos. The set was built to be diverse, once photos are distributed over seven categories, they were labeled as high or low quality. The problem of this photo set lies in the labeling process. Some important information is not shown, such as the exact number of votes for each category, the origin and background of the photographer, and the personal information of the voters. Besides this, a more precise ranking (instead of only classifying images as high/low quality) could be used for a more general use in enhancement and analysis algorithms. The other work, from Bhattacharya et al. [7], presented a smaller photo set (only 632 images). Other factors, such as the ones analyzed on the Luo et al. [94] approach, could not be evaluated, since the image set built by Bhattacharya et al. [7] could not be downloaded due to a Web server error. Thus, it might be considered that the image set is no longer available.

One might suggest that if it is possible to learn an expert opinion about a photo, it would be possible to analyze a photo. However this is surprisingly not always true. Since average photography consumers do not have training in what a good photo is, they often do not agree with advice given by experts. There are several other factors that might influence a photography user's opinion, such as photo effects and the event from which the photo was obtained, rather than photographic rules.

In conclusion, it was not possible to identify comparative studies involving different approaches, which considered publicly available photo datasets. This causes difficulties to reliably compare techniques when dealing with consumer photography. The use of image sets from photography contests has also its disadvantages since both photographers and voters may have professional skills or are highly interested in photography. This may lead to results that are not related to ordinary photography consumer preferences.

5.2 Validation

This section contains a discussion on validation approaches. Photography might be considered an art form [66]. There is no simple way of deciding whether a photo is aesthetically pleasant or not. However, it might be possible to identify some metrics that would help photo assessment, and that would be a step further in this area.

Another important aspect to be considered is how approaches were validated. Since the reviewed work is about photo enhancement and analysis, the results are usually images (in the case of photo enhancement algorithms) or abstract information, e.g., color/gray-scale maps, statistics, and scores. Both have a very high subjective component, although some metrics might be defined for obtaining a more objective analysis in a specific scenario.

Samsung EX1018
Page 13

Validation methods can be classified as subjective or objective. Subjective methods involve subjective experiments in which humans are asked to give their opinion on photos of a pre-defined test set with respect to a given attribute or criterion. A participant may give his/her opinion based on the following methods [98]:

- *Single-stimulus rating.* The participant will give a score to a photo or a group of photos. The score might be continuous (such as 0–10) or categorical (e.g., excellent, good, fair, bad, and poor). During the rating process, each photo is typically showed to the participant for a fixed presentation time (e.g., 3 s);
- *Double-stimulus rating.* While analogous to the single-stimulus rating, in double-stimulus trials a reference photo and a test photo are presented in random order, one after another, for a fixed presentation time (e.g., 3 s);
- *Forced-Choice.* The participant is forced to choose only one within a group of photos, according to a given criterion;
- *Pairwise similarity judgement.* Similar to forced-choice but, besides choosing one from a group of photos, the participant has also to indicate on a continuous scale how large the difference in quality is between the two photos; and
- *Indirect.* The participant does not directly give his/her opinion. The quality may be inferred by some measurement such as the time needed for the participant to choose a photo.

Other details and comparing methods can be found in the work of Mantiuk et al. [98], in which a comparison between the first four above-mentioned methods is given. The better method is usually the one with higher correlation between human and automatic labeling. It was shown, however, that for comparing IQA algorithms, in which differences between images might be small, the forced-choice pairwise comparison is the most accurate and time-efficient [98].

Besides the comparing method, there are also other factors have an influence in the experimental assessment, since such experiments involve humans. Some of the factors are:

- *Number of participants.* Once the opinion about the quality of a photo may vary from one person to another, it is important to have a large number of participants in order to identify features that are more significant in human analysis;
- *Used equipment.* When the experiment is conducted in an uncontrolled environment, the equipment used might harm the results (e.g., the calibration of the screen in a color experiment might produce a different opinion);
- *Knowledge in photography.* Experts evaluate photos in a different way than consumers do. As an example, professional photos might be considered good by both expert and consumer while a consumer photo might be considered good by a consumer but bad by experts;
- *Cultural diversity.* The style and subject of the photo might influence the judgement depending on the participant's background and origin; and
- *Number of photos.* The number of photos in the experiment is a factor as crucial as the number of participants. If, on the one hand, a great number of photos might better represent the diversity of the photos, on the other hand, it might reduce the number of volunteer participants, since it becomes a more laborious experiment.

Since, according to the literature review, there is no database which considers all those factors, most of the conclusions drawn from subjective experiments might be considered partially biased. Besides the drastic influence of such factors, there is no consensus on what are the ideal values for them. Thus, most papers present some questionable decisions on the validation step, such as the number of participants (e.g., three participants [108]), knowledge in photography (e.g., most participants are experts [73]), and the number of images used (e.g., only 34 photos to represent the analysis sample [49]).

On the other hand, objective metrics present a set of well-defined criteria, and proposals are evaluated based on those criteria. For instance, the best algorithm might be the one which has lower false-positive rates in a face detection scenario.

Objective methods are usually less expensive since they do not rely on the availability and classification coherence of human participants. However, there are some important features that are not yet well-assessed by computational algorithms, such as the global visual aspect of a photo. Even humans may disagree with a classification result, what may imply a harder subjective assessment. Both approaches (objective and subjective) are important, each in its specific application scenario.

As it can be seen in Tables 2 and 3, the methodology of the assessment widely differs in the reviewed work, with regards to the following aspects: (1) the assessment method, (2) the metric used for assessment, and (3) the source of the photo set.

Although the results reported in those tables were obtained with different algorithms and different goals, it is possible to conclude that most approaches have opted for subjective assessment when dealing with image enhancement and analysis. The reason is probably the lack of consensus on the image set to be used as ground-truth and the essentially subjective task of comparing images.

Samsung EX1018
Page 14

There is also a lack of clarity regarding the number of people used in the subjective experiments, their confidence with the labeling, and the methodology of the experiment.

6 Conclusions

This survey reviewed state-of-the-art methods for photo enhancement and analysis. For better understanding of this research area, a taxonomy was defined based on the related work. The main conclusions of this survey are discussed next:

- According to the conducted literature review, this is the first survey on consumer photographic enhancement and analysis techniques;
- The interest in algorithms for photo enhancement and analysis has been growing recently, based on the number of recent papers published in this area;
- There is not a consensus on a methodology for conducting subjective photo analysis experiments;
- Although the results were obtained with different algorithms and different goals, it is possible to conclude that most approaches have opted for a subjective assessment due to the lack of a public and labeled image set that might work as a ground-truth for an objective assessment, and due to the inherently subjective task of comparing images. Therefore, in this scenario, direct comparisons between existing approaches might be unfair;
- Some work that indicates the photo sources are not reproducible, since the photos used for testing are not clearly identified due mostly to copyright reasons or the great number of images;
- There is no consensus on the number of images to be used in the experiments; and
- There is a lack of clarity regarding the number of people used in the subjective experiments, their confidence with the labeling provided, and the assessment methodology.

Thus, it is possible to conclude that, although there has been recent growth in photo enhancement and analysis techniques, this is an area with large potential. Experimental assessment needs to be improved, and assessment methodologies are required as well in order to obtain strong conclusions about methods and results.

Acknowledgments The authors wish to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support of part of this research.

References

1. Achanta R, Süsstrunk, S (2009) Saliency detection for content-aware image resizing. In: Proceedings of the IEEE ICIP 2009. Piscataway, IEEE, pp 1005–1008
2. Anguelov, D, Lee KC, Gokturk SB, Sumengen B (2007) Contextual identity recognition in personal photo albums. In: Proceedings of the IEEE CVPR 2007. IEEE Computer Society, pp 1–7
3. Avidan S, Shamir A (2007) Seam carving for content-aware image resizing. ACM Trans Graphics 26(3):10.1-10.9
4. Banerjee S, Evans BL (2004) Unsupervised automation of photographic composition rules in digital still cameras. In: Proceedings of the SPIE Conference on sensors, color, cameras, and systems for digital photography, VI. pp 364–373
5. Benoit A, Caplier A, Durette B, Herault J (2010) Using human visual system modeling for bio-inspired low level image processing. Comput Vis Image Underst 114(7):758–773
6. Bertalmio M, Bugeau A, Caselles V, Sapiro G (2010) A comprehensive framework for image inpainting. IEEE Trans Image Process 19(10):2634–2645
7. Bhattacharya S, Sukthankar R, Shah M (2010) A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the ACM MM 2010, pp 271–280
8. Bissacco A, Yang M, Soatto S (2007) Fast human pose estimation using appearance and motion via multi-dimensional boosting regression. In: Proceedings of the IEEE CVPR 2007, pp 1–8
9. Boutell M, Luo J, Gray RT (2003) Sunset scene classification using simulated image recomposition. In: Proceedings of the IEEE ICME 2003, pp 37–40
10. Boyce W, Wilkie S (2013) Photosig. http://www.photosig.com. Accessed 31 January 2013
11. Bruneau P, Picarougne F, Gelgon M (2010) Interactive unsupervised classification and visualization for browsing an image collection. Pattern Recogn 43(2):485–493
12. Busselle M (1999) Better picture guide to photographing people. RotoVision, Hove
13. Byers Z, Dixon M, Goodier K, Grimm CM, Smart WD (2003) An autonomous robot photographer. In: Proceedings IEEE/RSJ IROS 2003, pp 2636–2641
14. Byers Z, Dixon M, Smart W, Grimm C (2004) Say cheese!: experiences with a robot photographer. AAAI Mag 25(3):37–46 (this is an invited paper that wraps up all of the other Lewis papers)
15. Cao L, Luo J, Kautz H, Huang T (2008) Annotating collections of photos using hierarchical event and scene models. In: Proceedings of the IEEE CVPR 2008, pp 1–8
16. Cao L, Luo J, Kautz H, Huang T (2009) Image annotation within the context of personal photo collections using hierarchical event and scene models. IEEE Trans Multiméd 11(2):208–219
17. Cavalcanti C, Gomes H, Veloso L, Carvalho J, Lima Jr O (2010) Automatic single person composition analysis. In: Skala V (ed) Proceedings of the WSCG 2010. UNION Agency-Science Press, Plzen, pp 229–236
18. Cavalcanti CSVC, Gomes H, Meireles R, Guerra W (2006) Towards automating photographic composition of people. In: Proceedings of the IASTED VIIP 2006. ACTA Press, Anaheim, pp 25–30
19. Celik T, Tjahjadi T (2010) Unsupervised colour image segmentation using dual-tree complex wavelet transform. Comput Vis Image Underst 114(7):813–826
20. Challenging technologies: dpchallenge a digital photography contest (2013) http://www.dpchallenge.com. Accessed 31 January 2013
21. Charrier C, Knoblauch K, Moorthy AK, Bovik AC, Maloney LT (2010) Comparison of image quality assessment algorithms on compressed images. In: Proceedings of the SPIE, Image Quality and System Performance VII, 2010. pp 75, 290B–1-75, 290B–11
22. Chartier S, Renaud P (2008) An online noise filter for eye-tracker data recorded in a virtual environment. In: Proceedings of the ACM ETRA 2008, pp 153–156

Samsung EX1018
Page 15

23. Chen H (2008) Note: Focal length and registration correction for building panorama from photographs. Comput Vis Image Underst 112(2):225–230
24. Chen Lq, Xie X, Fan X, Ma WY, Zhang Hj, Zhou HQ (2003) A visual attention model for adapting images on small displays. Multiméd Syst 9:353–364
25. Chen S, Cao L, Wang Y, Liu J, Tang X (2010) Image segmentation by map-ml estimations. IEEE Trans Image Process 19(9): 2254–2264
26. Chu WT, Lee YL, Yu JY (2009) Using context information and local feature points in face clustering for consumer photos. In: Proceedings of the IEEE ICASSP 2009, pp 1141–1144
27. Chu WT, Li CJ, Tseng SC (2011) Travelmedia: an intelligent management system for media captured in travel representation. J Vis Commun Image 22(1):93–104
28. Chu WT, Lin CH (2010) J Vis Commun Image Rep. Consumer photo management and browsing facilitated by near-duplicate detection with feature filtering 21(3):256–268
29. Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng YT (2009) Nus-wide: a real-world web image database from national university of singapore. In: Proceedings of ACM CIVR 2009, July 8–10
30. Cooper M, Foote J, Girgensohn A, Wilcox L (2005) Temporal event clustering for digital photo collections. ACM Trans Multiméd Comput Commun Appl 1(3):269–288
31. Corbis: Corbis image gallery (2001–2009). http://www.corbis.com. Accessed 31 January 2013
32. Corel Images: Corel images (2013) http://elib.cs.berkeley.edu/photos/corel/. Accessed 31 January 2013 (currently unavailable)
33. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
34. Creative Commons: Creative commons (2012) http://creativecommons.org/. Accessed 31 January 2013
35. Cunningham SJ, Masoodian M (2007) Identifying personal photo digital library features. In: Proceedings of the ACM/IEEE-CS JCDL 2007, pp 400–401
36. Daliri MR, Torre V (2009) Classification of silhouettes using contour fragments. Comput Vis Image Underst 113(9):1017–1025
37. Dao MS, Dang-Nguyen DT, De Natale FG (2011) Signature-image-based event analysis for personal photo albums. In: Proceedings of the ACM MM 2011, pp 1481–1484
38. Das M, Loui AC (2009) Event classification in personal image collections. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 1660–1663
39. Datta R, Joshi D, Li J, Wang JZ (2006) Studying aesthetics in photographic images using a computational approach. In: Proceedings of the ECCV 2006, pp 7–13
40. Datta R, Li J, Wang JZ (2008) Algorithmic inferencing of aesthetics and emotion in natural images: an exposition. In: Proceedings of the IEEE ICIP 2008, pp 105–108
41. Datta R, Wang JZ (2010) Acquine: aesthetic quality inference engine—real-time automatic rating of photo aesthetics. In: Proceedings of the ACM MIR 2010, pp 421–424
42. Destrero A, Mol C, Odone F, Verri A (2009) A regularized framework for feature selection in face detection and authentication. Int J Comput Vis 83(2):164–177
43. Duda RO, Stork DG, Hart PE (2000) Pattern classification and scene analysis. Part 1, Pattern classification, 2nd edn. Wiley, New York
44. Dunker P, Popp P, Cook R (2011) Content-aware auto-soundtracks for personal photo music slideshows. In: Proceedings of the IEEE ICME 2011, pp 1–5
45. Engelke U, Maeder AJ, Zepernick HJ (2009) On confidence and response times of human observers in subjective image quality assessment. In: Proceedings of the IEEE ICME 2009, pp 910–913
46. Ercegovac M, Lang T (1992) On-the-fly rounding (computing arithmetic). IEEE Trans Comput 41(12):1497–1503
47. Etchells D (2005) Canon expo 2005—a one-company trade show. http://www.imaging-resource.com/NEWS/1126887991.html. Accessed 31 January 2013
48. Fedorovskaya E, Neustaedter C, Hao W (2008) Image harmony for consumer images. In: Proceedings of the IEEE ICIP 2008, pp 121–124
49. Fedorovskaya E, Neustaedter C, Hao W (2008) Image harmony for consumer images. In: Proceedings of the IEEE ICIP 2008, pp 121–124. doi:10.1109/ICIP.2008.4711706
50. Flickr (2013) Flickr photo sharing. http://www.flickr.com/. Accessed 31 January 2013
51. Flickr (2013) Mirflickr-25000. http://www.flickr.com/photos/tags/. Accessed 31 January 2013
52. Foggia P, Percannella G, Sansone C, Vento M (2008) Int J Pattern Recogn Artif Intell. A graph-based algorithm for cluster detection 22(5):843–860
53. Fujita H, Arikawa M (2007) Creating animation with personal photo collections and map for storytelling. In: Proceedings of the ACM EATIS 2007. ACM, New York, pp 1:1–1:8
54. Gallagher A, Chen T (2008) Clothing cosegmentation for recognizing people. In: Proceedings of the IEEE CVPR 2008, pp 1–8
55. Gallagher AC, Chen T (2007) Using group prior to identify people in consumer images. In: Proceedings of the IEEE CVPR 2007, vol 0. IEEE Computer Society, pp 1–8
56. Gallagher AC, Chen T (2009) Finding rows of people in group images. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 602–6058
57. Golder S (2008) Measuring social networks with digital photograph collections. In: Proceedings of the ACM HT 2008, pp 43–48
58. Google: Picasa (2013) http://picasa.google.com/. Accessed 31 January 2013
59. Gorelick L, Basri R (2009) Shape based detection and top–down delineation using image segments. Int J Comput Vis 83(3): 211–232
60. Greenspun P (2013) Photo.net photography community. http://photo.net. Accessed 31 January 2013
61. Gupta A, Chen F, Kimber D, Davis LS (2008) Context and observation driven latent variable model for human pose estimation. In: Proceedings of the IEEE CVPR 2008, pp 1–8
62. Haddad Z, Beghdadi A, Serir A, Mokraoui A (2010) Image quality assessment based on wave atoms transform. In: Proceedings of the IEEE ICIP 2010, pp 305–308
63. Han HS, Kim DO, Park RH (2009) Structural information-based image quality assessment using lu factorization. IEEE Trans Consum Electron 55(1):165–171
64. Han J, Awad G, Sutherland A (2009) Automatic skin segmentation and tracking in sign language recognition. IET-CV 3(1):24–35
65. Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, Englewood Cliffs
66. Hedgecoe J (2009) New manual of ohotography. Dorling Kindersley, New York
67. Hoíng NV, Gouet-Brunet V, Rukoz M, Manouvrier M (2010) Embedding spatial information into image content description for scene retrieval. Pattern Recogn 43(9):3013–3024
68. Hsu SH, Jumpertz S, Cubaud P (2008) A tangible interface for browsing digital photo collections. In: Proceedings of the ACM TEI 2008, pp 31–32
69. Ji H, Liu C (2008) Motion blur identification from image gradients. In: Proceedings of the IEEE CVPR 2007, vol 0, IEEE Computer Society, pp 1–8
70. Jiang H, Martin D (2008) Global pose estimation using non-tree models. In: Proceedings of the IEEE CVPR 2008, pp 1–8

Samsung EX1018
Page 16

71. Jiang W, Loui A, Cerosaletti C (2010) Automatic aesthetic value assessment in photographic images. In: Proceedings of the IEEE ICME 2010, pp 920–925
72. Joshi N, Matusik W, Adelson EH, Kriegman DJ (2010) Personal photo enhancement using example images. ACM Trans Graphics 29(2):1–15
73. Ke Y, Tang X, Jing F (2006) The design of high-level features for photo quality assessment. In: Proceedings of the IEEE CVPR 2006, pp 419–426
74. Khan SS, Vogel D (2012) Evaluating visual aesthetics in photographic portraiture. In: Proceedings of the CAe 2012. Eurographics Association, pp 55–62
75. Kim HN, Saddik AE, Jung JG (2012) Leveraging personal photos to inferring friendships in social network services. Expert Syst Appl 39(8):6955–6966
76. Kruppa H, Bauer MA, Schiele B (2002) Skin patch detection in real-world images. In: Proceedings of the the 24th DAGM Symposium on Pattern Recognition. Springer LNCS, pp 109–117
77. Lee C, Schramm MT, Boutin M, Allebach JP (2009) An algorithm for automatic skin smoothing in digital portraits. In: Proceedings of the IEEE ICIP 2009. IEEE Press, New York, pp 3113–3116
78. Lee S, Kim K, Kim JY, Kim M, Yoo HJ (2010) Familiarity based unified visual attention model for fast and robust object recognition. Pattern Recogn 43(3):1116–1128
79. Levin A, Weiss Y (2009) Learning to combine bottom–up and top–down segmentation. Int J Comput Vis 81(1):105–118
80. Li C, Gallagher AC, Loui AC, Chen T (2010) Aesthetic quality assessment of consumer photos with faces. In: Proceedings of the IEEE ICIP 2010, pp 3221–3224
81. Li C, Loui AC, Chen T (2010) Towards aesthetics: a photo quality assessment and photo selection system. In: Proceedings of the ACM MM 2010, pp 827–830
82. Li X, Ling H (2009) Learning based thumbnail cropping. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 558–561
83. Li Z, Luo H, Fan J (2009) Incorporating camera metadata for attended region detection and consumer photo classification. In: Proceedings of the ACM MM 2009, pp 517–520
84. Li Z, Zhou X, Huang TS (2009) Spatial gaussian mixture model for gender recognition. In: Proceedings of the IEEE ICIP 2009. IEEE Press, New York, pp 45–48
85. Liao WH (2009) A framework for attention-based personal photo manager. In: Proceedings of the the IEEE SMC 2009. IEEE Press, New York, pp 2128–2132
86. Lim SH, Lin Q, Petruszka A (2010) Automatic creation of face composite images for consumer applications. In: Proceedings of the IEEE ICASSP 2010, pp 1642–1645
87. Liu L, Chen R, Wolf L, Cohen-Or D (2010) Optimizing photo composition. In: Proceedings of the Eurographics, vol 29, pp 469–478
88. Liu R, Li Z, Jia J (2008) Image partial blur detection and classification. In: Proceedings of the IEEE CVPR 2007, vol 0. IEEE Computer Society, Los Alamitos, pp 1–8
89. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum HY (2011) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
90. Liu Y, Xu D, Tsang IW, Luo J (2011) Textual query of personal photos facilitated by large-scale web data. IEEE Trans Pattern Anal Mach Intell 33(5):1022–1036
91. Loui A, Wood M, Scalise A, Birkelund J (2008) Multidimensional image value assessment and rating for automated albuming and retrieval. In: Proceedings of the IEEE ICIP 2008, pp 97–100
92. Loui AC, Wood MD, Scalise A, Birkelund J (2008) Multidimensional image value assessment and rating for automated albuming and retrieval. In: Proceedings of the IEEE ICIP 2008, pp 97–100
93. Lu F, Yang X, Zhang R, Yu S (2009) Image classification based on pyramid histogram of topics. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 398–401
94. Luo W, Wang X, Tang X (2011) Content-based photo quality assessment. In: Proceedings of the IEEE ICCV 2011, vol. 0. IEEE Computer Society, Los Alamitos, pp 2206–2213
95. Luo Y, Tang X (2008) Photo and video quality evaluation: focusing on the subject. In: Proceedings of the ECCV 2008. Springer, Heidelberg, pp 386–399
96. Lux M, Kogler M, del Fabro M (2010) Why did you take this photo: a study on user intentions in digital photo productions. In: Proceedings of the ACM SAPMIA 2010, pp 41–44
97. Maik V, Paik D, Lim J, Park K, Paik J (2010) Hierarchical pose classification based on human physiology for behaviour analysis. IET-CV 4(1):12–24
98. Mantiuk RK, Tomaszewska A, Mantiuk R (2012) Comparison of four subjective methods for image quality assessment. Comput Graphics Forum 31(8):2478–2491
99. Marshall B (2010) Taking the tags with you: Digital photograph provenance. In: Proceedings of the IEEE symposium on data, privacy, and E-Commerce 2010. IEEE Computer Society, Los Alamitos, pp 72–77
100. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the ICCV 2001, vol 2, pp 416–423
101. Mele K, Suc D, Maver J (2009) Local probabilistic descriptors for image categorisation. IET-CV 3(1):8–23
102. Microsoft Corporation: Bing maps (2012) http://www.bing.com/maps/. Accessed 31 January 2013
103. Miller AD, Edwards WK (2007) Give and take: a study of consumer photo-sharing culture and practice. In: Proceedings of the ACM SIGCHI 2007, pp 347–356
104. Moorthy AK, Obrador P, Oliver N (2010) Towards computational models of the visual aesthetic appeal of consumer videos. In: Proceedings of the ECCV 2010. Springer, Berlin/Heidelberg, pp 1–14
105. Mu Y, Yan S, Liu Y, Huang T, Zhou B (2008) Discriminative local binary patterns for human detection in personal album. In: Proceedings of the IEEE CVPR 2008, vol 0. New York, IEEE Computer Society, pp 1–8
106. Mucientes M, Bugarín A (2010) People detection through quantified fuzzy temporal rules. Pattern Recogn 43(4):1441–1453
107. Nikon Corporation: Nikon d90 advanced function (2008). http://chsvimg.nikon.com/products/imaging/lineup/d90/en/advanced-function/. Accessed 31 January 2013
108. Obrador P (2008) Region based image appeal metric for consumer photos. In: Proceedings of the IEEE Workshop on multimedia signal 2008, pp 696–701
109. Obrador P, Moroney N (2009) Automatic image selection by means of a hierarchical scalable collection representation. In: Proceedings of the SPIE visual communications and image processing, San Jose, vol 7257, pp 0W.1–0W.12
110. Obrador P, de Oliveira R, Oliver N (2010) Supporting personal photo storytelling for social albums. In: Proceedings of the ACM MM 2010, pp 561–570
111. Obrador P, Schmidt-Hackenberg L, Oliver N (2010) The role of image composition in image aesthetics. In: Proceedings of the IEEE ICIP 2010, pp 3185–3188
112. O'Hare N, Lee H, Cooray S, Gurrin C, Jones G, Malobabic J, O'Connor N, Smeaton AF, Uscilowski, B (2006) Mediassist: Using content-based analysis and context to manage personal photo collections. In: Proceedings of the CIVR 2006, vol 4071. Springer, Heidelberg, pp 529–532

Samsung EX1018
Page 17

113. O'Hare N, Smeaton AF (2009) Context-aware person identification in personal photo collections. IEEE Trans Multiméd 11(2):220–228
114. Oliveira CJS, Araújo AdeA, Severiano CA Jr, Gomes DR (2002) Classifying images collected on the World Wide Web. In: Proceedings of the SIBGRAPI 2002, IEEE Computer Society Press, Fortaleza, pp 327–334
115. Orbanz P, Buhmann JM (2008) Nonparametric bayesian image segmentation. Int J Comput Visi 77(1–3):25–45
116. Paisitkriangkrai S, Shen C, Zhang J (2008) Performance evaluation of local features in human classification and detection. IET-CV 2(4):236–246
117. Pang Y, Hao Q, Yuan Y, Hu T, Cai R, Zhang L (2011) Summarizing tourist destinations by mining user-generated travelogues and photos. Comput Vis Image Underst 115(3):352–363
118. Park HJ, Har DH (2011) Subjective image quality assessment based on objective image quality measurement factors. IEEE Trans Consumer Electron 57(3):1176–1184
119. Peres M (2007) Focal encyclopedia of photography: digital imaging, theory and applications, history, and science. Elsevier Science Inc./Focal Press, Boston
120. Pierrard JS, Vetter T (2007) Skin detail analysis for face recognition. In: Proceedings of the IEEE CVPR 2007, pp 1–8
121. Presti LL, Cascia ML (2012) An on-line learning method for face association in personal photo collection. Image Vis Comput 30 (4–5):306–316
122. Qi GJ, Hua XS, Rui Y, Tang J, Zhang HJ (2008) Two-dimensional active learning for image classification. In: Proceedings of the IEEE CVPR 2008, pp 1–8
123. Qin AK, Clausi DA (2010) Multivariate image segmentation using semantic region growing with adaptive edge penalty. IEEE Trans Image Process 19(8):2157–2170
124. Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106
125. Rahman M, Gamadia M, Kehtarnavaz N (2008) Real-time face-based auto-focus for digital still and cell-phone cameras. In: Proceedings of the IEEE SSIAI 2008. IEEE Computer Society, Los Alamitos, pp 177–180
126. Redi JA, Heynderickx I (2012) Image integrity and aesthetics: towards a more encompassing definition of visual quality. In: Proceedings of the SPIE human vision and electronic imaging XVII 2012, vol 8291. SPIE, San Jose, pp 15.1–15.10
127. Ren T, Liu Y, Wu G (2009) Image retargeting based on global energy optimization. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 406–409
128. Ren X, Fowlkes CC, Malik J (2008) Learning probabilistic models for contour completion in natural images. Int J Comput Vis 77 (1–3):47–63
129. Rousson M, Paragios N (2008) Prior knowledge, level set representations & visual grouping. Int J Comput Vis 76(3):231–243
130. Russell SJ, Norvig P (2009) Artificial intelligence: a modern approach, 3rd edn. Prentice Hall, New Delhi
131. Ryoo MS, Aggarwal JK (2009) Semantic representation and recognition of continued and recursive human activities. Int J Comput Vis 82(1):1–24
132. Sandnes F (2010) Unsupervised and fast continent classification of digital image collections using time. In: Proceedings of the ICSSE 2010, pp 516–520
133. Santella A, Agrawala M, Decarlo D, Salesin D, Cohen M (2006) Proceedings of the gaze-based interaction for semi-automatic photo cropping. In: Proceedings of the ACM SIGCHI 2006. ACM Press, New York, pp 771–780
134. Savakis AE, Etz SP, Loui ACP (2000) Evaluation of image appeal in consumer photography. In: Proceedings of the SPIE human vision and electronic imaging V, vol 3959. SPIE, San Jose, pp 111–120
135. Schindler G, Krishnamurthy P, Lublinerman R, Liu Y, Dellaert F (2008) Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In: Proceedings of the IEEE CVPR 2008, pp 208–219
136. Schmugge SJ, Jayaram S, Shin MC, Tsap LV (2007) Objective evaluation of approaches of skin detection using roc analysis. Comput Vis Image Underst 108(1–2):41–51
137. Serrano N, Savakis A, Luo J (2002) A computationally efficient approach to indoor/outdoor scene classification. In: Proceedings of the IEEE ICPR 2002. IEEE Computer Society, Los Alamitos, pp 146–149
138. Setlur V, Takagi S, Raskar R, Gleicher M, Gooch B (2005) Automatic image retargeting. In: Proceedings of the ACM MUM 2005. ACM Press, New York, pp 59–68
139. Shen CT, Liu JC, Shih SW, Hong JS (2009) Towards intelligent photo composition-automatic detection of unintentional dissection lines in environmental portrait photos. Expert Syst Appl 36(5):9024–9030
140. Singla P, Kautz H, Gallagher A (2008) Discovery of social relationships in consumer photo collections using markov logic. In: Proceedings of the IEEE CVPR 2008 Workshops, pp 1–7
141. Sinha P (2011) Summarization of archived and shared personal photo collections. In: Proceedings of the ACM WWW 2011, pp 421–426
142. Snavely N, Seitz SM, Szeliski R (2008) Modeling the world from internet photo collections. Int J Comput Vis 80(2):189–210
143. Sony Corporation: Sony party-shot automatic photographer (2009). http://store.sony.com/webapp/wcs/stores/servlet/ProductDisplay?catalogId=10551&storeId=10151&langId=-1&partNumber=IPTDS1. Accessed 31 January 2013
144. Stein A, Stepleton T, Hebert M (2008) Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In: Proceedings of the IEEE CVPR 2008, pp 1–8
145. Su HH, Chen TW, Kao CC, Hsu WH, Chien SY (2011) Scenic photo quality assessment with bag of aesthetics-preserving features. In: Proceedings of the ACM MM 2011, pp 1213–1216
146. Suh B, Bederson BB (2007) Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. Interact Comput 19(4):524–544
147. Suh B, Ling H, Bederson BB, Jacobs DW (2003) Automatic thumbnail cropping and its effectiveness. In: Proceedings of the ACM UIST 2003. ACM Press, New york, pp 95–104
148. Tang F, Gao Y (2009) Fast near duplicate detection for personal image collections. In: Proceedings of the ACM MM 2009, pp 701–704
149. Tian A, Zhang X, Tretter DR (2011) Content-aware photo-on-photo composition for consumer photos. In: Proceedings of the ACM MM 2011, pp 1549–1552
150. Tómasson G, Sigurp'orsson H, Jónsson B, Amsaleg L (2011) Photocube: effective and efficient multi-dimensional browsing of personal photo collections. In: Proceedings of the ACM ICMR 2011, pp 70:1–70:2
151. Tong H, Li M, Zhang H, Zhang C (2004) Blur detection for digital images using wavelet transform. In: Proceedings of the IEEE ICME 2004, pp 17–20
152. Tong H, Li M, Zhang HJ, He J, Zhang C (2004) Classification of digital photos taken by photographers or home users. In: Proceedings of the Pacific Rim Conference on Multimedia. Springer, Heidelberg, pp 198–205
153. Tran C, Wijnhoven R, de With P (2011) Text detection in personal image collections. In: Proceedings of the IEEE ICCE 2011, pp 85–86
154. Tsao WK, Lee AJT, Liu YH, Chang TW, Lin HH (2010) A data mining approach to face detection. Pattern Recogn 43(3): 1039–1049

Samsung EX1018
Page 18

155. Tsay KE, Wu YL, Hor MK, Tang CY (2009) Personal photo organizer based on automated annotation framework. International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp 507–510
156. Valle E, Cord M, Philipp-Foliguet S, Gorisse D (2010) Indexing personal image collections: a flexible, scalable solution. IEEE Trans Consumer Electron 56(3):1167–1175
157. Vogel J, Schwaninger A, Wallraven C, Bülthoff HH (2007) Categorization of natural scenes: Local versus global information and the role of color. ACM Trans Appl Percept 4(3):19.1–19.21
158. Wan D, Zhou J (2008) Stereo vision using two ptz cameras. Comput Vis Image Underst 112(2):184–194
159. Wang H, Oliensis J (2010) Generalizing edge detection to contour detection for image segmentation. Comput Vis Image Underst 114(7):731–744
160. Wang J, Jia Y, Hua XS, Zhang C, Quan L (2008) Normalized tree partitioning for image segmentation. In: Proceeings of the IEEE CVPR 2008, vol 0. IEEE Computer Society, Los Alamitos, pp 1–8
161. Wang J, Zhu S, Gong Y (2009) Resolution-invariant image representation for content-based zooming. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 918–921
162. Wang P, Ji Q (2007) Multi-view face and eye detection using discriminant features. Comput Vis Image Underst 105(2):99–111
163. Wang XF, Huang DS, Xu H (2010) An efficient local chan-vese model for image segmentation. Pattern Recogn 43(3):603–618
164. Wang Y, Huang Q, Gao W (2009) Pornographic image detection based on multilevel representation. IJPRAI 23(8):1633–1655
165. Wichmann FA, Drewes J, Rosas P, Gegenfurtner KR (2010) Animal detection in natural scenes: critical features revisited. J Vis 10(4):6.1–27
166. Xie S, Shan S, Chen X, Chen J (2010) Fusing local patterns of gabor magnitude and phase for face recognition. IEEE Trans Image Process 19(5):1349–1361
167. Xie ZX, Wang ZF (2010) Color image quality assessment based on image quality parameters perceived by human vision system. In: Proceedings of the ICMT 2010, pp 1–4
168. Xin H, Ai H, Chao H (2011) Tretter D Human head-shoulder segmentation. In: Proceedings of the IEEE FG 2011, pp 227–232
169. Xu S, Ye X, Wu Y, Giron F, Leveque JL, Querleux B (2008) Automatic skin decomposition based on single image. Comput Vis Image Underst 110(1):1–6
170. Xu Z, Sun J (2010) Image inpainting by patch propagation using patch sparsity. IEEE Trans on Image Process 19(5):1153–1165
171. Yan S, Wang H, Liu J, Tang X, Huang TS (2010) Misalignment-robust face recognition. IEEE Trans Image Process 19(4): 1087–1096
172. Yanagawa A, Loui AC, Luo J, Chang SF, Ellis D, Jiang W, Kennedy L, Lee K (2008) Kodak consumer video benchmark data set: concept definition and annotation. Columbia University, Technical report
173. Yang Y, Xu D, Nie F, Yan S, Zhuang Y (2010) Image clustering using local discriminant models and global integration. IEEE Trans Image Process 19(10):2761–2773
174. Yanulevskaya V, van Gemert J, Roth K, Herbold A, Sebe N, Geusebroek J (2008) Emotional valence categorization using holistic image features. In: Proceedings of the IEEE ICIP 2008, pp 101–104
175. Yao L, Suryanarayan P, Qiao M, Wang JZ, Li J (2012) Oscar: on-site composition and aesthetics feedback through exemplars for photographers. Int J Comput Vis 96(3):353–383
176. Yeh CH, Ho YC, Barsky BA, Ouhyoung M (2010) Personalized photograph ranking and selection system. In: Proceedings of the ACM MM 2010, pp 211–220
177. Yeh CH, Ng WS, Barsky BA, Ouhyoung M (2009) An esthetics rule-based ranking system for amateur photos. In: Proceedings of the ACM SIGGRAPH 2009, pp 24:1–24:1
178. Yi Y, Yu X, Wang L, Yang Z (2008) Image quality assessment based on structural distortion and image definition. In: Proceedings of the international conference on computer science and software engineering 2008(6):253–256
179. Yin W, Luo J, Chen CW (2010) Semantic adaptation of consumer photo for mobile device access. In: Proceedimgs of the ISCAS 2010, pp 1173–1176
180. Ying Z, Guangyao L, Xiehua S, Xinmin Z (2009) Geometric active contours without re-initialization for image segmentation. Pattern Recogn 42(9):1970–1976
181. Yu Z, Au OC, Zou R, Yu W, Tian J (2010) An adaptive unsupervised approach toward pixel clustering and color image segmentation. Pattern Recogn 43(5):1889–1906
182. Zeng G, Gool LV (2008) Multi-label image segmentation via point-wise repetition. In: Proceedings of the IEEE CVPR 2008, vol 0. IEEE Computer Society, Los Alamitos, pp 1–8
183. Zeng YC (2009) Automatic local contrast enhancement using adaptive histogram adjustment. In: Proceedings of the IEEE ICME 2009. IEEE Press, New York, pp 1318–1321
184. Zha ZJ, Hua XS, Mei T, Wang J, Qi GJ, Wang Z (2008) Joint multi-label multi-instance learning for image classification. In: Proceedings of the IEEE CVPR 2008, vol 0. IEEE Computer Society, Los Alamitos, pp 1–8
185. Zhang H, Fritts JE, Goldman SA (2008) Image segmentation evaluation: a survey of unsupervised methods. Comput Vis Image Underst 110(2):260–280
186. Zhang M, Zhang L, Sun Y, Feng L, Ma W (2005) Auto Cropping for Digital Photographs. In: Proceedings of the IEEE ICME 2005, pp 438–441
187. Zhang T, Chao H, Willis C, Tretter D (2010) Consumer image retrieval by estimating relation tree from family photo collections. In: Proceedings of the ACM CIVR 2010, pp 143–150
188. Zhou C, Lin S (2007) Removal of image artifacts due to sensor dust. In: Proceedings of the IEEE CVPR 2007, vol 0. IEEE Computer Society, Los Alamitos, pp 1–8

Samsung EX1018
Page 19