Journal of Visual Communication and Image Representation 54 (2018) 39–50

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Color uniformity descriptor: An efficient contextual color representation for image indexing and retrieval

Carolina Reta[a,*], Jose A. Cantoral-Ceballos[b], Ismael Solis-Moreno[c], Jesus A. Gonzalez[d], Rogelio Alvarez-Vargas[b], Nery Delgadillo-Checa[b]

[a] Department of IT, Control, and Electronics, CONACYT-CIATEQ A. C., Av. Diesel Nacional #1 Ciudad Sahagun, Hidalgo 43990, Mexico
[b] Department of IT, Control, and Electronics, CIATEQ A. C., Av. Manantiales #23-A, Parque Industrial Bernardo Quintana, El Marques, Queretaro 76246, Mexico
[c] Mexico Software Lab, IBM, Carretera al Castillo km. 2.2, El Salto, Jalisco 45686, Mexico
[d] Department of Computer Science, National Institute for Astrophysics, Optics, and Electronics, Luis Enrique Erro #1, Tonantzintla, Puebla 72840, Mexico

ARTICLE INFO

Keywords:
Image retrieval
Image representation
Contextual features
Color uniformity descriptor
Lab color space

ABSTRACT

Color is a rich source of visual information for the effective characterization of image content. The recognition of texture or shape elements in images is strongly associated with the analysis of the image color layout. This paper presents a contextual color descriptor designed especially to be applied to CBIR tasks in heterogeneous image databases. The proposed color uniformity descriptor (CUD) clusters perceptually similar image color regions according to the uniformity analysis of their neighbor pixels. CUD produces vast color image details with a thin histogram, whilst preserving the balance between uniqueness and robustness. CUD is computationally efficient and can achieve high precision and throughput rates when used in CBIR. Experimental results show that CUD performs comparably against local features and multiple features state-of-the-art approaches that require more complex data manipulation. Results demonstrate that CUD provides strong image discrimination even in the presence of significant content variation.

1. Introduction

The fast-growing amount of digital images together with the need to explore them effectively, encourages the research and development of efficient techniques for image storage, indexing, and retrieval [1]. Applications of image storage and retrieval cover a wide variety of fields including social networking, media, advertising, art, architecture, education, medicine, biometry, industry, engineering, geology, astronomy, and remote sensing [2]. In the case of image retrieval, there are two general frameworks: text-based and content-based. Text-based image retrieval (TBIR) is widely used in commercial search engines, such as Google and Bing. TBIR systems use a set of keywords to index images with similar metadata descriptions. Their main disadvantage is that they require human intervention to annotate images manually. This task is unfeasible in heterogeneous and large-scale image collections due to errors introduced by the subjectivity of human perception and by the significant amount of participants required. In contrast, content-based image retrieval (CBIR) systems are completely automatic. They use any visual information extracted from the image itself (e.g. color, edges, texture, shape, or local features) to index images with similar image content. The representation of an image by its content is a challenging task since the most relevant information has to be extracted from the image, so that it is well represented in a significantly lower dimensionality space [1,3]. A variety of image-content descriptors based on different visual features has been proposed in the literature. However, some state-of-the-art approaches are not suitable for real-time CBIR in heterogeneous and large-scale datasets due to expensive training phases and/or sampling strategies to extract visual features [4].

In this paper, we propose the color uniformity descriptor (CUD) and its corresponding image retrieval system. The CUD is an efficient contextual color descriptor designed specifically for real-time CBIR tasks in heterogeneous image databases. The CUD characterizes the image content by analyzing pixel color and the effect of its neighborhood context. The CIE $L^*a^*b^*$ color representation of the image is used to obtain the CUD. This fact allows summarizing significant uniformity color features of perceptually similar color regions into a compact histogram whilst maintaining uniqueness and robustness balance, e.g. 210 bins. The CUD calculation is computationally efficient, achieving high precision and throughput rates when used in CBIR – as

* Corresponding author.
E-mail addresses: carolina.reta@ciateq.mx (C. Reta), jose.cantoral@ciateq.mx (J.A. Cantoral-Ceballos), ismael@mx1.ibm.com (I. Solis-Moreno), jagonzalez@inaoep.mx (J.A. Gonzalez), ralvarez@ciateq.mx (R. Alvarez-Vargas), nery.delgadillo@ciateq.mx (N. Delgadillo-Checa).

https://doi.org/10.1016/j.jvcir.2018.04.009
Received 2 March 2017; Received in revised form 8 March 2018; Accepted 21 April 2018
Available online 26 April 2018

experimentation conducted over four different standard image collections reveals. The experimentation demonstrates that CUD achieves competitive results that, in some cases, exceed the performance of notable state-of-the-art descriptors with much higher time complexity. It is noteworthy that CUD provides a good trade-off between generality and particularity allowing strong image discrimination in the presence of significant content variation. This is further validated with the introduction of image distractors. The robustness, compactness, and efficiency of CUD make it suitable for use in heterogeneous and large-scale image collections.

The remaining sections of the paper are organized as follows. Section 2 discusses the state-of-the-art work on image-content descriptors that can be used in CBIR. Section 3 explains the detail of CUD and the proposed image retrieval system. Section 4 describes the conducted experimentation to evaluate CUD in standard image collections. Also, it presents a comparison of the achieved results with other state-of-the-art approaches. Finally, Section 5 presents the conclusions and describes the future work.

2. Review of image-content descriptors

Image content is represented by the analysis of any kind of visual information derived from the image, e. g. color, edges, texture, shape, and local features. Visual features can be extracted from entire images describing global visual features or from key points or salient patches describing local visual features [4–6]. The most notable efforts in both global and local features-based descriptors are presented in the following subsections.

2.1. Descriptors based on color features

Color histograms are the most commonly used color descriptors in CBIR [1,7–10]. A color histogram describes the global color distribution in an image represented by a specific color space, such as RGB, YCbCr, HSV, CIE XYZ, CIE $L^*a^*b^*$, CIE $L^*u^*v^*$, YIQ, YUV, HMMD, and Opponent [7–9,11]. For instance, the MPEG-7 scalable color descriptor (SCD) defines a global color histogram encoded by a Haar transform in the HSV color space [8,11,12]. Color histograms are fast to compute and are robust against small changes in camera, e.g. viewpoint, object's position, size, and pose variations, as well as object's partial occlusions. In [9], color invariant histograms, derived from the RGB, HSV, and opponent color spaces, are proposed to make color histograms robust against changes in illumination, shadows, and highlights. However, color histograms have a limited discriminating effect since they do not consider the spatial organization of colors. Thus, images with very different content can be represented by color histograms with similar color distribution. This problem becomes especially critical for CBIR in large-scale image collections [13]. Therefore, other descriptors have been designed to integrate spatial information into histograms, including the MPEG-7 descriptors: dominant color descriptor (DCD), color layout descriptor (CLD), and color structure descriptor (CSD) [8,11,12], as well as the color coherence vector (CCV) [10], and color correlograms [13]. DCD gives the global as well as the local spatial distribution of the salient colors in the image [8,12]. CLD succinctly describes the spatial distribution of colors in the YCbCr color space [8,11]. CSD obtains the spatial distribution of colors in the HMMD color space contained within a structural element [8,11]. In [14], authors combine the characteristics of DCD and CSD into a single descriptor called Dominant Color Structure Descriptor (DCSD). The color coherence vector (CCV) [10] partitions each histogram bin into two types, coherent if pixels of that color are members of similarly-colored regions, and incoherent otherwise. In [15], CCV in the $CIL^*a^*b^*$ color space (Lab-CCV) was proposed to represent the spatial organization of color pixel intensities in a perceptually uniform manner. CCV provides good retrieval results in images which have either mostly uniform color or mostly texture regions [15,16]. The color auto-correlogram [13] combines the spatial correlation of color regions and the global distribution of local spatial correlation of colors. Although it can achieve good retrieval results, it is computationally expensive due to its high dimensionality [16].

2.2. Descriptors based on edge and texture features

Since they consider the context of the pixel in its local neighborhood, they can deal with noise and small lighting variations. The most common edge features-based descriptors are edges histograms and histograms of oriented gradient (HOG). While edge histograms [8,17] are obtained by counting the directionality of the brightness gradients within the whole image, HOG [18] features are obtained by computing normalized local histograms of gradient directions over a grid of small spatial regions. Edge histograms can be fast to compute since their computational complexity depends on the method used to extract edge features. Hence, they are a good option for CBIR systems. However, the use of HOG features in CBIR systems is not recommended, since they also require an expensive training stage to obtain significant visual features that can discriminate among a variety of image content. Notable texture-based descriptors defined in the MPEG-7 standard are the texture browsing descriptor (TBD), the homogeneous texture descriptor (HTD), and the local edge histogram descriptor (EHD) [8,12,17]. TBD characterizes the regularity, directionality, and coarseness of textures [8,17]. HTD provides the characterization of homogeneous regions by computing local spatial-frequency statistics of the texture [8,17]. EHD captures the spatial distribution of edges, which is used as a texture signature in images with non-homogeneous texture properties [8,12]. Except for the TBD descriptor, which uses too little statistical measures to describe the image content, the HTD and EHD descriptors provide a good compact representation to describe the image content in a global way. Their use is recommended for CBIR in image collections with high interclass variability because they are better generalizing the image content than capturing its particularities. A good texture descriptor is LBP (local binary patterns), a grayscale invariant texture model derived from a local neighborhood. Due to its robustness to extract local patterns from images, such as edges, spots, curves, and flat areas [1,19], it has been successfully applied in some CBIR systems [1,15,20,21].

2.3. Descriptors based on local features (LF)

They emphasize salient local regions in the image. Notable LF detectors are Harris, Shi-Tomasi, FAST, SIFT, SURF, BRISK, BRIEF, and ORB [4,22–24]. However, their direct use for CBIR is impractical because hundreds, or even thousands of LF are extracted per image [4]. LF-based CBIR approaches usually employ subsampling strategies to extract LF, then represented in Bag of Visual Words (BOVW) [4,25]. Image retrieval for BOVW is performed using inverted indexes [4,20,26–28]. In [4], sixteen SIMPLE descriptors that use four blob-like patch generators (SIFT, SURF, Random, and Gaussian Random) for sampling, and four global descriptors (SCD, CLD, EHD, and CEDD – presented Sect. 2.4) for describing the patches, are proposed. The results show that the combination of both random generated patches produce better performance than those generated by SIFT and SURF for CBIR tasks, in almost all combinations. This is explained by the fact that, contrasting object recognition problems, in CBIR it is not always critical to achieve one-to-one matching in image sections. The co-indexing algorithm [20] allows editing inverted indexes by incorporating both LF and semantic attributes to enhance its discriminative ability. The improving Bag of Features [26] enhances inverted indexes by refining the matching of visual words using a Hamming embedding (HE) signature and by removing matching descriptors that are not consistent in angle and scale. An alternative that uses an asymmetric HE approach is proposed in [27]. The coupled-binary embed algorithm [29] refines visual matching via the embedding of multiple binary features. Alternative models to BOVW include LF histograms [24], LF signatures [24],