Volume 6, No. 6; December 2018

Advances in Image And Video Processing

ISSN: 2054-7412

Society for Science and Education - United Kingdom

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 6, ISSUE 6
ISSN 2054-7412

Improved Stereo Image Dehazing Approach

[1]Yu-Chen Wu, [2]Chin-Chen Chang, [3]Der-Lor Way, [4]Zen-Chung Shih

[1,4]Institute of Multimedia Engineering, National Chiao Tung University, Hsinchu, Taiwan;
[2]Department of Computer Science and Information Engineering, National United University, Miaoli Taiwan;
[3]Department of NewMedia Art, Taipei National University of Arts, Taipei, Taiwan;
ccchang@nuu.edu.tw

ABSTRACT

Numerous image dehazing algorithms have been studied intensively. However, most dehazing algorithms operate on single images. These algorithms produce inconsistent results if they are used to dehaze stereo images iteratively. In this paper, we present a novel dehazing approach for stereo images based on cross bilateral filtering. In this approach, we simultaneously estimate scene depth and dehaze the stereo images. The proposed approach is based on the observation of depth cues in the stereo images. Depth cues are mainly used to avoid inconsistent results, and the cross bilateral filter is used to preserve shape details. The results demonstrate that the proposed approach can deliver superior results to those of previously published methods.

Keywords: Dehazing; Stereo Images; Image Restoration; Depth Estimation.

1 Introduction

If a scene is photographed under inclement weather conditions, such as fog or haze, the quality of the captured images would be degraded. This is not a direct process of degradation; the light is attenuated as it moves toward the camera. The light entering the camera is blended with ambient light reflected from the atmospheric particles. Therefore, images not only are degraded but also lose contrast and color fidelity. Recently, three-dimensional (3D) image technology has been developed intensively. Investment in cutting-edge image production and services, such as Google glasses, Google cardboard, virtual reality, hand-held devices, and wearable systems, has also increased greatly.

However, most recently developed systems are not suitable for outdoor use. One of the reasons is that the produced stereo images are associated with poor image contrast, which causes image features to be less distinctive and confuses stereo matching. In particular, stereo images can be used for creating or enhancing the illusion of depth by stereopsis; this is notably demonstrated by human binocular vision. Stereoscopic visual information from the left and right eyes is combined in the brain to create the perception of 3D depth. Thus, it is necessary to restore images when dealing with outdoor stereo images.

Despite the high demand for 3D outdoor applications or services, few visibility restoration algorithms have been proposed for foggy stereo images. One naive solution is to apply a basic dehazing algorithm to stereo images iteratively. However, most dehazing algorithms are designed to run on a single image. The iterative dehazing of stereo images might cause inconsistent results between the stereo images [9]. In this paper, we study stereo vision and defogging problems jointly, and present an algorithm that simultaneously estimates scene depth and defogs the input stereo images.

DOI: 10.14738/aivp.66.5497
Publication Date: 25th November 2018
URL: http://dx.doi.org/10.14738/aivp.66.5497

The proposed approach is based on the observation of depth cues in stereo images. The fog transmission can be directly computed from depth. In this approach, we use the dark channel prior (DCP) technique [7] to improve the transmission results. However, this causes other challenges, including halo effects and false contours, to arise. Thus, we enhance our transmission calculation by combining it with the method of Shiau et al. [12] and with a bilateral filter [16]. The bilateral filter can preserve edges to avoid false contours and the method of Shiau et al. [12] can prevent halo effects. Finally, our approach can avoid inconsistent results between a pair of stereo images based on a depth map. The proposed approach can achieve a better result than a naive method that applies a dehazing algorithm to stereo images iteratively.

The remainder of this paper is organized as follows: In Section 2, we review related works about single image dehazing and stereo image dehazing. Section 3 describes our proposed method and the details of each procedure. Section 4 discusses the results. Section 5 draws conclusions and suggests directions for future research.

2 Related Works

Image dehazing is challenging because hazy scenes often suffer from low contract and limited visibility. Therefore, numerous single image dehazing algorithms have been proposed. A typical single image dehazing algorithm attempts to recover a single image from a hazy image. Because a hazy image only conveys ambiguous information about the depicted scene, some single-image dehazing methods require additional data. Kopf et al. [8] used a user registration process to produce depth information and dehaze an image in the context of depth. Some single-image dehazing methods are based on color restoration. Tan's method [13] maximizes local contrast and can obtain airlight, which is the light scattered by particles. Fattal [4] obtained color lines in an image and used them to dehaze the image. Fattal also proposed another method [3] that entails combining the transmission and shading models, and these models were applied to dehaze images. Fattal assumed that, locally, transmission and shading are statistically uncorrelated.

Some other single-image dehazing methods were based on filters. Gibson and Nguyen [5] estimated the amount of fog in an image with a local Wiener filter. They used the theory of Retinex [10], [17] for single image dehazing. Multiscale Retinex with color restoration (MSRCR) can improve the visual rendering of an image even when the lighting conditions are not adequate. Menget et al. [10] introduced a boundary constraint on a transmission function. Tarel and Hautiere [14] introduced a method based on a median filter and combined with an edge preserving filter. He et al. [7] introduced a DCP that exploits the minimum intensity of all color channels in a local window to indicate the level of haziness. Shiau et al. [12] improved the method of He et al. [7] by applying a weighted technique that automatically determines the possible atmospheric lights and mixes these candidates to refine the atmospheric light.

Several dehazing algorithms have been proposed for stereo images. Li et al. [9] presented a method for video defogging and stereo reconstruction. They obtained the similarity between video frames and constructed a depth map according to this similarity. They improved the photo-consistency term to explicitly model the appearance change due to the scattering effect. However, Carr and Hartley [1] proposed an approach to enhance a single image by compensating for fog effects, assuming that the depth and atmospheric conditions are known. Kopf et al. [8] demonstrated that even a rough depth information can also significantly improve the dehazing performance. Moreover, Li et al. [9] concluded that a stereo-image dehazing algorithm based on the constraints of a depth map could obtain a

favorable result. The constraints of depth maps can counteract some inconsistencies between stereo images.

3 The Proposed Approach

The proposed approach initially computes the transmission and disparity. The transmission is computed on the basis of the weighted dark channel method [12]. Simultaneously, the scattering coefficient and atmospheric light vector are computed. Next the fine transmission is calculated based on the cross bilateral filter with transmission and depth maps. Finally, the fine transmission can be transformed to a fine depth map directly or used for dehazing.

3.1 Stereo Disparity Map

The depth or disparity map plays a major role because a fine depth map can produce an excellent dehazing result [8]. The depth map of a pair of stereo images can be computed by two processes, namely camera calibration and disparity creation. Camera calibration can be used to undistort the image. In this step we use the method of Zhang [21] for camera calibration. Because we focus on stereo image dehazing, we calibrate the left and right cameras. Therefore, we undistort the left and right images iteratively.

For the disparity map, we use Yang's method [18]. This approach combines nonlocal cost aggregation and a minimum spanning tree. It can produce excellent results in regions with low textures and low complexity. This method can retain most of the shape details. This is especially suitable for a set of hazy stereo images. Yang's method is suitable for parallel programming [19].

3.2 Dark Channel Weighted Transmission

The depth or disparity map plays a major role because a fine depth map can produce an excellent dehazing result [8]. The depth map of a pair of stereo images can be computed by two processes, namely camera calibration and disparity creation. Camera calibration can be used to undistort the image. In this step we use the method of Zhang [21] for camera calibration. Because we focus on stereo image dehazing, we calibrate the left and right cameras. Therefore, we undistort the left and right images iteratively.

A transmission is useful when recovering a hazy image. In this step, we recover the transmission through the DCP technique [7]. However, the DCP result is not of sufficient quality. We combine the DCP technique with the weighting function of Shiau et al. [12]. For a hazy image, the observed light is absorbed and scattered by turbid mediums such as atmospheric particles or raindrops. In the context of this observation, a simple atmospheric scattering model [11] can be used to represent the formation of a hazy image. This model is defined as Equations (1) and (2). This model is also used in several dehazing algorithms [1, 3-5, 7-10, 12-15, 17, 20].

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

$$t(x) = e^{-\beta d(x)} \tag{2}$$

A guidance hazy image I can be expressed by Equation (1), where J is the clear image, A is the global atmospheric light and t is the transmission. When the atmosphere is homogeneous, the transmission t can be expressed as Equation (2), where β is the scattering coefficient and $d(x)$ is the distance from the scene to the camera center. Equation (2) indicates that the scene radiance attenuates exponentially with the depth. Therefore, we can recover a fine depth map if we can recover the transmission.

For each haze-free image, He et al. [7] found that in most of the nonsky patches, at least one color channel has some pixels whose intensities are remarkably low and close to zero. Similarly, the minimum

intensity in such a patch is close to zero. Accordingly, He et al. [7] defined such observations as dark channels; these can be formally represented as given in Equation (3):

$$\min_{y\in\Omega}\left(\min_{c} J_c(x)\right) = 0 \tag{3}$$

where c indicates the color channel, which is one of red, green, or blue. Hence, J_c is the color channel of any arbitrary image J and Ω is a local patch centered at x. A dark channel is the outcome of two minimum operators.

With the dark channel and the assumption that the transmission in a local patch is constant, Equation (1) can be written as Equation (4). We denote this transmission as $\tilde{t}$, and this transmission is constant in the patch.

$$\min_{y\in\Omega}\left(\min_{c}\frac{I_c(x)}{A_c}\right) = \min_{y\in\Omega}\left(\min_{c}\frac{J_c(x)}{A_c}\right)\tilde{t}(x) + (1-\tilde{t}(x)) \tag{4}$$

By substituting Equation (3) into Equation (4), we can obtain a transmission recovery method as expressed in Equation (5).

$$\tilde{t}(x) = 1 - \min_{y\in\Omega}\left(\min_{c}\frac{I_c(x)}{A_c}\right) \tag{5}$$

The transmission obtained directly from the dark channel is not of adequate quality, especially in the sky region or night image. This is because the estimation of a dark channel is based on patches. Therefore, a distortion problem tends to arise for any dehazing result at any boundary between sky and nonsky regions.

He et al. [7] solved the problem with soft matting. However, that solution is not generally suitable because soft matting is a time-consuming method and might generate inconsistent results for stereo images. Shiau et al. [12] proposed an idea of weighted transmission. The main problem of distortion of the dark channel method is the assumption required for Equation (4). It is assumed that the transmission in a local patch is constant. However, this assumption is inconsistent with the actual result, particularly within any patches between sky and nonsky regions. For the dark channel method, the local minimum or local maximum in such a patch would affect the patch transmission balance. The weighted transmission can avoid such circumstances.

Similarly, we define a difference function in Equation (6). This difference function calculates the ratio of the current pixel dark channel relative to the patch dark channel.

$$diff(x) = \frac{\min_{c}\{I_c(x)\}}{\min_{y\in\Omega}\left\{\min_{c}\{I_c(y)\}\right\}} \tag{6}$$

The weighting function is defined as Equation (7), where θ is a nonzero positive constant.

$$W_\Omega = \frac{e^{-\theta[diff(x)-1]}}{diff(x)} \tag{7}$$

This weighting function concludes that if there exists a pixel with a significant difference in dark channel relative to the patch, the weight of that pixel must be larger than the weights of other pixels in the same patch. With the aforementioned weighting function, we can rewrite the transmission in Equation (4). Finally, the new transmission is stated as Equation (8).

$$\tilde{t}(x) = 1 - \left((1-W_\Omega)\times\min_{c}\left\{\frac{I^c(x)}{A^c}\right\} + W_\Omega \times \min_{y\in\Omega}\left\{\min_{c}\left\{\frac{I^c(x)}{A^c}\right\}\right\}\right) \tag{8}$$

3.3 Cross Bilateral Filtering

We use a bilateral filter [2] to jointly combine the depth map and the transmission map. A bilateral filter is a nonlinear, edge-preserving and noise-reducing smoothing filter for images [22]. The intensity value of each pixel is calculated based on a weighting function. The weight is based on a Gaussian distribution; the weight depends on the Euclidean distances between pixels and radiometric differences [16].

The bilateral filter in the proposed approach is used to preserve the edges and restrict the constraint of depth cues. However, a stereo depth map does not work well for large distances [6]. Nevertheless, the transmission contains a depth ordering and a smoothness prior probability. Therefore, in our approach, we use the depth map as a set of radiometric differences to restrict the adjacent pixels' intensities accordingly.

In the proposed approach, the cross bilateral filter is defined as Equation (9).

$$t_p = \frac{1}{W_p} \sum_{q \in \Omega_p} G_s(\|p - q\|) G_r(|D_p - D_q|)\, \tilde{t}_q \tag{9}$$

where $\tilde{t}$ indicates the transmission image created from dark channel weighted transmission, D is the disparity map, Ω_p is the center patch at pixel p, and W_p is a normalized term, defined by Equation (10).

$$W_p = \sum_{q \in \Omega_p} G_s(\|p - q\|) G_r(|D_p - D_q|) \tag{10}$$

The spatial similarity G_s and the range similarity G_r are defined in Equations (11) and (12), where σ^2 and ϱ^2 are nonzero positive constants.

$$G_s(\|p - q\|) = \exp\left\{-\frac{\|p-q\|}{2\sigma^2}\right\} \tag{11}$$

$$G_r(|D_p - D_q|) = \exp\left\{-\frac{|D_p - D_q|}{2\varrho^2}\right\} \tag{12}$$

The result t_p is the final fine transmission. We can recover the fine depth map directly using Equation (2).

4 Results

The proposed approach was run on a PC with a 3.33 GHz Core™ i7 CPU and 8.0 GB of memory. The patch size Ω was used in the dark channel weighted transmission and cross bilateral filter. For relatively small patches, the dehazed images were oversaturated. However, for relatively large patches, the halo effects were stronger along the edges between the sky and nonsky regions.

We used UNITY to simulate fog effects. Figures 1-3 illustrate the results of the fog simulation with an exponential function. For the method of He et al. [7], the dehazed images appeared overly saturated. For the method of Shiau et al. [12], the colors of distant objects seemed unnatural.

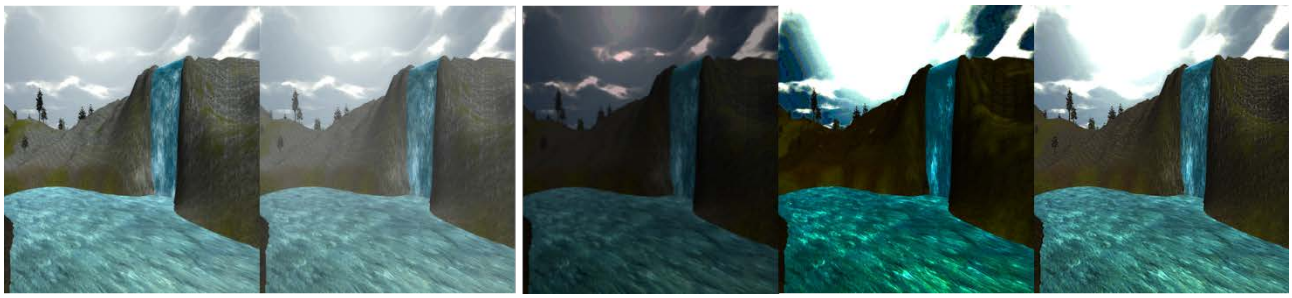


Figure 1. Waterfall: (a) original image, (b) haze image, (c) image dehazed by the method of He et al. [7], (d) image dehazed by the method of Shiau et al. [12], and (e) image dehazed by our approach.

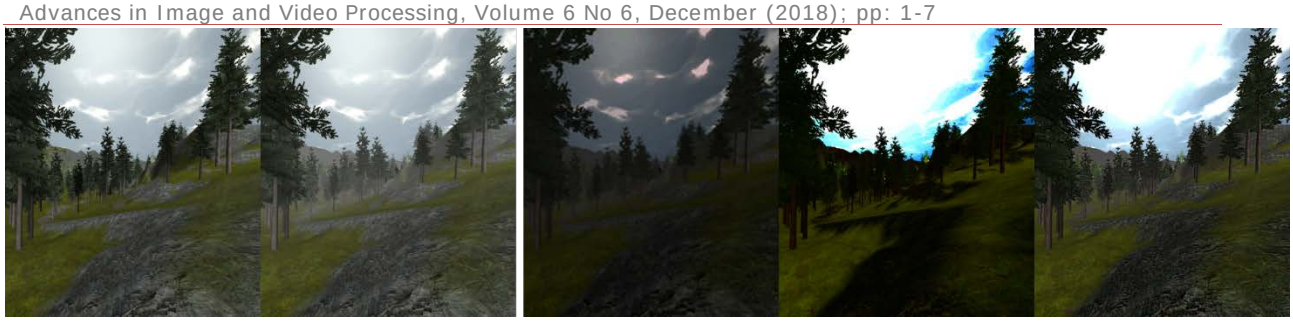


Figure 2. Terrain: (a) original image, (b) haze image, (c) image dehazed by the method of He et al. [7], (d) image dehazed by the method of Shiau et al. [12], and (e) image dehazed by our approach.

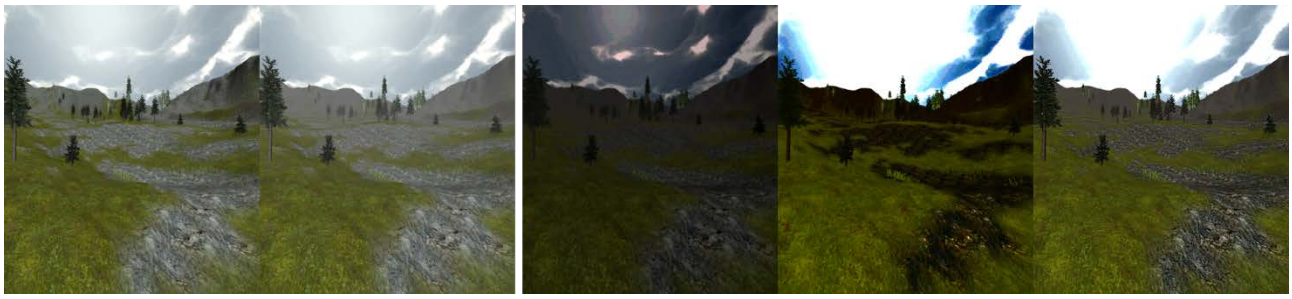


Figure 3. Rock road: (a) original image, (b) haze image, (c) image dehazed by the method of He et al. [7], (d) image dehazed by the method of Shiau et al. [12], and (e) image dehazed by our approach.

5 Conclusion

In this paper, we have proposed a dehazing approach for stereo images. Our approach is based on depth cues and avoids inconsistent results. A cross bilateral filter can constrain the transmission and preserve shape details. The results prove that our approach can obtain high quality dehazed results. The results also demonstrate that it is not proper to dehaze stereo images iteratively. Dehazing methods that incorporate depth information should be considered when dehazing stereo images.

Because our approach requires a disparity map, the quality of the disparity map influences the quality of our approach. The proposed approach may fail for objects with similar colors. We will address this problem for future research.

REFERENCES

[1] P. Carr, R. Hartley, "Improved Single Image Dehazing Using Geometry", Digital Image Computing: Techniques and Applications (DICTA), 103–110, 2009.

[2] H. Cho, H. Lee, H. Kang, S. Lee, "Bilateral Texture Filtering", ACM Transactions on Graphics (TOG), 33(4), 1281–1288, 2014.

[3] R. Fattal, "Single Image Dehazing", ACM Transactions on Graphics (TOG), 27(3), 7–23, 2008.

[4] R. Fattal, "Dehazing Using Color-Lines", ACM Transactions on Graphics (TOG), 34(1), 1–14, 2014.

[5] K. Gibson, T. Nguyen, "Fast Single Image Fog Removal Using the Adaptive Wiener Filter", Proceedings of the International Conference on Image Processing (ICIP), 714–718, 2013.

[6] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge: Cambridge University Press, 2004.

[7] K. He, J. Sun, X. Tang, “Single Image Haze Removal Using Dark Channel Prior”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 33(12), 2341–2353, 2011.

[8] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, “Deep Photo: Model-Based Photograph Enhancement and Viewing”, ACM SIGGRAPH Asia, 2008.

[9] Z. Li, P. Tan, R.T. Tan, D. Zou, “Simultaneous Video Defogging and Stereo Reconstruction”, Computer Vision and Pattern Recognition (CVPR), 4988–4997, 2015.

[10] G. Meng, Y. Wang, J. Duan, X. Xiang, C. Pan, “Efficient Image Dehazing with Boundary Constraint and Contextual Regularization”, International Conference on Computer Vision (ICCV), 617–624, 2013.

[11] W.E.K. Middleton, Vision Through the Atmosphere, Toronto: University of Toronto Press, 1952.

[12] Y.H. Shiau, P.Y. Chen, H.Y. Yang, C.H. Chen, S.S. Wang, “Weighted Haze Removal Method with Halo Prevention”, Journal of Visual Communication and Image Representation, 25(2), 445–453, 2014.

[13] R.T. Tan, “Visibility in Bad Weather from A Single Image”, Computer Vision and Pattern Recognition (CVPR), 1–8, 2008.

[14] J.P. Tarel, H. Hautiere, “Fast Visibility Restoration from a Single Color or Gray Level Image”, International Conference on Computer Vision (ICCV), 2201–2208, 2009.

[15] J.P. Tarel, H. Hautiere, “Improved Visibility of Road Scene Images under Heterogeneous Fog”, Proceedings of IEEE Intelligent Vehicle Symposium (IV), 478–485. 2010.

[16] C. Tomasi, R. Manduchi, “Bilateral Filtering For Gray and Color Images”, International Conference on Computer Vision (ICCV), 839–846, 1998.

[17] J.B. Wang, N. He, K. Lu, “A New Single Image Dehazing Method with MSRCR Algorithm”, Proceedings of the International Conference on Internet Multimedia Computing and Service (ICIMCS), 2015.

[18] Q.X. Yang, “A Non-Local Cost Aggregation Method For Stereo Matching”, Computer Vision and Pattern Recognition (CVPR), 1402–1409, 2012.

[19] Q.X. Yang, N. Ahuja, K.H. Tan, “Constant Time Median and Bilateral Filtering”, International Journal of Computer Vision, 112(3), 307–318, 2015.

[20] Y.C. Wu, C.C. Chang, D.L. Way, Z.C. Shih, “A Novel Dehazing Approach for Stereo Images,” in Proceedings of the International Workshop on Advanced Image Technology 2017 (IWAIT 2017), Penang, Malaysia, January 2017.

[21] Z. Zhang, “A Flexible New Technique for Camera Calibration”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 22(11), 1330–1334, 2000.

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 6, ISSUE 6
ISSN 2054-7412

Flower Image Retrieval Using Color and Shape Features and Multiple Distance Functions

[1]Khadiza Sultana Happy, [2]Zannatul Azme , [3,*]Mohammad Farhad Bulbul
[1,2,3]Department of Mathematics, Jessore University of Science and Technology; Bangladesh
Khadijasultana007@gmail.com; zannatulazme1510@gmail.com; farhad@just.edu.bd

ABSTRACT

It is noticeable that when flower images are observed, flower's color signifies people and often people ignore small and isolated sectors of the image. So for retrieving flower images color feature of the images is very important. Different flowers have also different shapes. So for differing flower images from another flower images shape feature is also a significant feature. In this paper, we proposed a method of retrieving flower images using their color and shape feature. We used HSV color histogram, color moments and color auto correlogram descriptors for describing color feature and HOG descriptor for describing shape feature. We calculate the similarities between the query image and the database images by employing a set of distance functions. We evaluated our result for our own dataset of 15 categories flower images.

Keywords: Flower image retrieval; HSV histogram; Color moments; HOG feature; L2 distance function.

1 Introduction

In ancient time people used the walls to draw some pictures to tell some information about their life. At the beginning of the twentieth century, an ever-growing number of digital images are being gathered with the introduction of digital photography. Nowadays, images play a vital role in various fields such as medical, education, and entertainment. By using computers, various techniques help us for capturing, processing, storage, and transmitting the images. Using World Wide Web users are facilitated to access data from anywhere and can utilize the images in any fields [1]. Although for this reason, the quantity of data becomes larger and larger, it will be unusable unless there are active methods to access them.

Images collected from several fields are stored in image databases. And the images for each field have certain common features. In many fields such as medical, engineering, educational, sports, criminal etc. images are used. For example, X-ray images are used in medical fields for diagnoses and research purpose. For finding the suspicious people face recognition is used in the criminal field. If we want to search an image by text or image description, the result will not accurate and it will time consuming. So it is not effectual to use text or image descriptors to search for an image in a huge image database. A new method called content-based image retrieval is used to retrieve images from the database to overwhelm this problem [2].

Content-based image retrieval (CBIR) is a process to retrieve images based on their visual features like color, texture, and shape [3]. The conception of using content-based image retrieval was first used by Kato to describe his experiments for retrieving images from a database using color and shape features.

DOI: 10.14738/aivp.66.5618
Publication Date: 23rd December 2018
URL: http://dx.doi.org/10.14738/aivp.66.5618

Nowadays CBIR methods have become a dependable tool for many image database applications. There are several benefits of CBIR techniques.

Flower image retrieval method is a type of content-based image retrieval method and it is very effectual for computer-aided flower species recognition. With the availability of digital photo devices, the collection of flower images can be enormous in size, containing hundreds, thousands or even millions of flower images in a database. There are many difficulties to retrieve flower images for being its complex backgrounds, for variants of brightness and intensity under diverse natural illuminations, shadow effects of the neighboring, different camera viewpoints, different dimensions, and different resolution, etc. As a result, the retrieval of flower images is a challenging problem in content-based image retrieval.

Many researchers have been researched on flower image retrieval system. In 2008, Maria-Elena Nilsback and Andrew Zisserman [3] make known to a flower dataset with 103 class. They compute four different features namely the local shape/texture, the shape of the boundary, the overall spatial distribution of petals, and the color of the flowers. They merged the features using a multiple kernel framework with a classifier called SVM. The weights for each class were learnt using the technique of Varma and Ray [4], which has succeeded state of the art performance on another huge dataset, such as Caltech 101/256. Their dataset had a comparable contest in the number of classes. Their results displayed that learning the optimum kernel combination of multiple features vastly improves the performance, from 55.1% for the best single feature to 72.8% for the combination of all features.

A few years later, Lin Li, Yu Qiao [5] presented a new idea for flower image retrieval. They created by using of category attributes to construct middle-level image representation. Image retrieval has been utilized by low-level features. They utilized category attributes, i.e. daisy, buttercup or iris to construct a semantic representation of flower images. They trained a linear SVM based on low-level visual features, containing the appearance of color, texture and shape for each category attribute. These classifiers are regarded as attribute features. So that, they construct a new mid-level visual presentation - category attribute - for the query and images in the dataset. They were evaluated on 17 category flower dataset. Finally, they took the result of captures flower category attribute that was more effective and had lower dimension. So they did not need to rely on costly and inefficiently similarity measurement between low-level feature vectors.

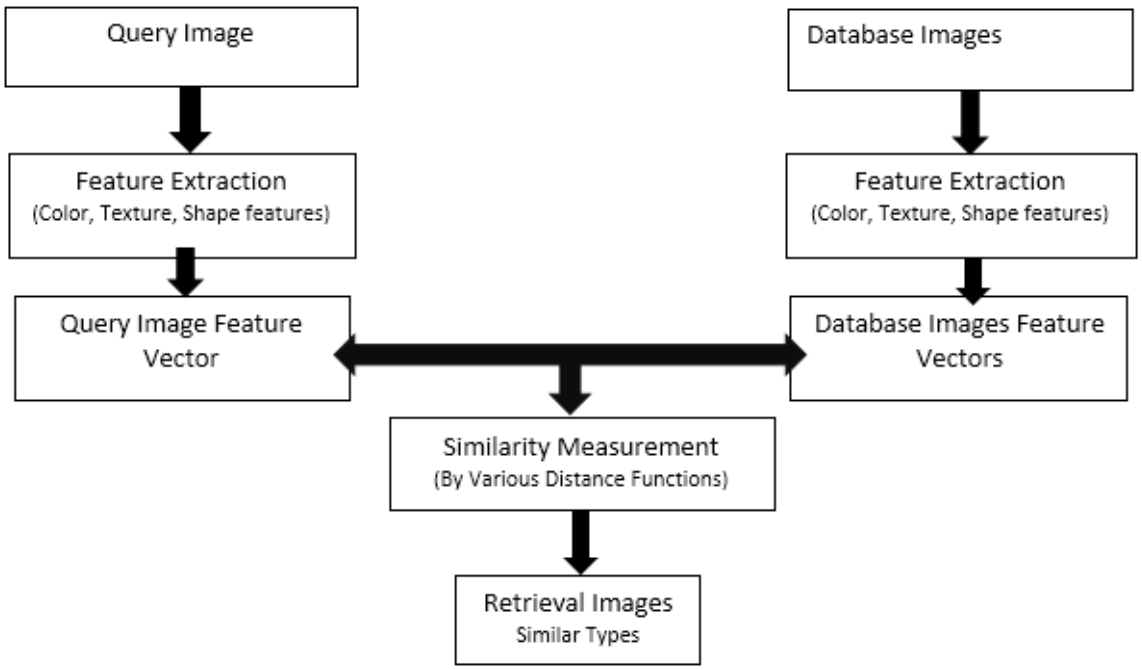
After that, Y H Sharath Kumara*, D S Gurub [6] propose a model based on a query sketch for representation and indexing of flower images for the purpose of retrieving flowers of interest. They used a subset of flower sketches having 20 users and for each user 5 flower sketches, totally about 100 sketches. Their thought was an effective way too rapid up the retrieval procedure is to design an indexing scheme initially for retrieval, which results in the reduction of the number of flowers considered for searching by the retrieval algorithm. They represented each flower by shape descriptors of Scale Invariant Feature Transform (SIFT), Histogram of Gradients (HOG) and Edge Orientation Histograms (EOH). Their Experimentations were conducted on Category- 127dataset and Flower Sketch dataset (their own flower data set of 127 classes of flowers, consists of 13169 flower images) to assess the advantage of using indexing technology. Their experimentation ensures that the combination of all the features descriptors achieves a good accuracy with indexing approach.

2 Proposed Method

2.1 Architectural Overview

In this part, we will discuss about our method for retrieving flower images. Firstly, extracting and calculating the features we create feature vectors for the query image and database images. Then using some distance functions through the feature vectors we find out the similarity between the query image and the database images. Figure 1 represents an architectural overview of the retrieval system.

Figure 1: User interface of the retrieval system



2.2 Feature Extraction

Using the contents of images, CBIR technique represent the image. From the value of the image pixels CBIR systems extract features (color, texture, and shape). The size of these features are smaller than the size of the images and they are stored in a database called feature database. In the feature database each image is represented by a combine representation of its contents (color, texture, shape, and spatial information) in the form of a fixed length real-valued multi-component feature vectors or signature. This is called offline feature extraction [7].

In our work we use descriptors for color features and Histogram of oriented gradients (HOG) for shape feature to describe a flower image. This section discusses those features briefly.

2.2.1 Color Feature

Color is the most spontaneous feature of an image and to describe colors HSV histogram, color moments and color auto correlogram are used in our method. Histogram techniques have the advantages of rapidness, lower demand of memory space. Color features are the most significant elements facilitating human to recognize images. For categorizing images, powerful information can be provide by color features and they are used for image retrieval.

(i)HSV histogram: Anyone with a monitor has probably heard of the RGB color space (Jacci Howard bear). HSV is a variation of an RGB color space, and its components and colorimetry are comparative to the RGB color space from which it was obtained. We may have noticed **HSV** in the color picker of our graphics software when we deal with commercial printers. HSV are common cylindrical-coordinate representations of points in an RGB color model. The HSV color model describes in terms of hue, saturation and value (brightness).Hue corresponds right away to the opinion of hue in the color Basics section. It is the dominate color which is perceived by an observer. Saturation corresponds directly to the concepts of tint in the color .Basics section except that full saturation produces on tint while zero saturation produces white a shade of gray, or black. Value corresponds directly to the concepts of intensity in the color Basics section. The HSV color wheel is sometimes illustrated as a cone or cylinder. But it always with these three components. It is used when choosing colors for draw or ink as HSV better represents how people relate to colors than dose the RGB color space. Selecting an HSV color starts with picking one of the available hues, which is how most humans anticipate to color, and then conforming the shade and brightness value. The benefit of HSV is that each of its quality corresponds right away to the basic color thought, which makes it conceptually instinctive. RGB to HSV conversion formula [8] is shown as follows:

The *R*,*G*,*B* values are distributed by 255 to change the range from 0..255 to 0..1:

R' = *R*/255

G' = *G*/255

B' = *B*/255

Cmax = max(*R*', *G*', *B*')

Cmin = min(*R*', *G*', *B*')

Δ = *Cmax* - *Cmin*

Hue calculation:

$$H = \begin{cases} 0^\circ & \Delta = 0 \\ 60^\circ \times \left(\frac{G'-B'}{\Delta} mod 6\right), & C_{max} = R' \\ 60^\circ \times \left(\frac{B'-R'}{\Delta} + 2\right) & , C_{max} = G' \\ 60^\circ \times \left(\frac{R'-G'}{\Delta} + 4\right), & C_{max} = B' \end{cases} \quad (1)$$

Saturation calculation:

$$S = \begin{cases} 0\ , & C_{max} = 0 \\ \frac{\Delta}{C_{max}}, & C_{max} \neq 0 \end{cases} \quad (2)$$

Value calculation:

$$V = C_{max} \quad (3)$$

Where,

H(Hue)=Dominant color(spectral), S(saturation)=Amount of white, V(values)=Brightness

Figure 2: Example of HSV images of flower images

Original Images ⟶ HSV Images

(ii) Color moments: Color moments have been successfully used in content based image retrieval system. It is used for extraction of color feature. It can be used differentiate images based on their features of color are measured by color moments. The number of color distribution in an image in the identical way are measured by it. It is measures central moments uniquely relating a probability distribution. It is mainly used for color indexing purposes as features in image retrieval applications. Probability distribution are characterized by a number of unique moments. Color moments are scaling and rotation invariant. It can be computed for any color model. Stricker and Orengo [9] use three central moments of an image's color distribution. They are Mean, Standard deviation and Skewness. A color can be described by 3 or more values. It is generally the case that only the first three color moments are used as features in color image retrieval system that is HSV hypothesis of Hue, saturation and value. Three color moments are computed per channel. An image therefore is characterized by 9 moments-3 moments for each color channels. Calculating color moments is done in the same way as computing moments of a probability distribution. Now we will identify the m-th color direction at the n-th image pixel as p_{mn} . The three color moments can then be defined as,

Mean: It can be calculated by using the following formula,

$$E_m = \sum_{n=1}^{N} \frac{1}{N} p_{mn} \tag{4}$$

Where,

N= the number of pixels in the images

p_{mn}= the value of the n-th pixel of the image at the m-th color channel.

Mean can be assumed as the middling color value in the image.

Standard Deviation: It can be calculated by using the following formula,

$$\sigma_m = \sqrt{\left(\frac{1}{N}\sum_{n=1}^{N}(p_{mn} - E_m)^2\right)} \tag{5}$$

Where

E_m = The mean value or first color moment for the m-th color channel.

The standard deviation is the square root of the adaptation of the distribution.

Skewness: It can be calculated by using the following formula,

$$s_m = \sqrt[3]{(\frac{1}{N}\sum_{n=1}^{N}(p_{mn} - E_m)^3)} \quad (6)$$

Skewness can be perceive as a measure of the dimension of asymmetry in the distribution.

(iii) Color auto correlogram: Color is an important visual feature. We have been successfully used of color auto correlogram algorithm in color based image retrieval as a feature descriptor. It is used for extraction of color feature. Correlogram can be conserved as a table indexed by pairs of colors (m,n) where d-th entry shows the probability of finding a pixel n from pixel m at distance d. Whereas a color auto-correlogram can be conserved as a table index of color in the same pixel at distance d.Hence auto-correlogram captures spatial correlation into same colors only [10].

Let [D] denote a set of D given distances d1? dD. Then the correlogram of the image M is defined for level pair $g_m; g_n$ at a distance d is given by,

$$\gamma_{g_m,g_n}^{d}(M) = P_{s_{p1\in Mg_m,p2\in M}}\left[p2 \in M_{g_n} \parallel p1 - p2 = d \parallel\right] \quad (7)$$

This gives the probability that given any pixel p1 of level g_m, a pixel p2 at a distance d in certain direction from the given pixel p1 is of level g_m.

Color auto correlogram captures the spatial correlation of same levels only [11]:

$$\alpha_g^d(M) = \gamma_{g_m g_n}^{d}(M) \quad (8)$$

It gives the probability that pixels p1 and p2, d away from each other, are of the same levelg_m.

2.2.1 Shape Feature

The physical structure of the objects, or the geometric shapes present in the image. Objects may be recognized from their outline. Shape is very powerful feature. Shape feature must present some essential properties such as identifiability, translation, rotation and scale invariance, noise resistance, occultation invariance, statistically independent, reliability etc. [12]

(i) Histogram of oriented gradients (HOG): The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision. The HOG feature are widely use of image processing for the purpose of object detection. The detector can be distributed into HOG feature extraction and SVM classification. HOG feature is used for extraction of shape feature. HOG decomposes an image into small squared cells. It counts a histogram of oriented gradients in every cell. It also normalizes the result using a block-wise Patten then return a descriptor for each cell. HOG is easy to express the rough shape of the object. It is robust to variations in geometry. But HOG is not supported into rotation and scale changes [13]. HOG feature used in algorithm such as color image is converted to grayscale, luminance gradient is calculated at each pixel.

When calculate a HOG descriptor, we need to first calculate the horizontal and vertical gradients. Because we want to calculate the histogram of gradients. The horizontal and vertical gradients by filtering the kernels as the table below, then calculated the histogram of gradients,

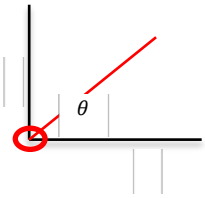
-1	0	1

-1
0
1

We know that, gradient basically is defined as a measure of change in an image function. Gradient of an image function is defined as mathematically given as follows,

Let, $F(X,Y)$is the image function. Where, X represents the horizontal direction and Y represents the vertical direction.

Centered: $F' = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$



$$\text{Gradient of an image} \quad \nabla F = \left[\frac{\delta F}{\delta x}, \frac{\delta F}{\delta y}\right] \tag{9}$$

Where,

$$F_x = \frac{\delta F}{\delta x} = F(x+1) - F(x)\,,$$

Which is denoted by differential of the function image F w.r.to X.

$$F_y = \frac{\delta F}{\delta y} = F(y+1) - F(y)$$

Which is denoted by differential of the function image F w.r.to y.

This is the way to measure of change in image function F in both $x\ and\ y$.

Now we can find the magnitude, direction and normalization of gradient using the following formula,

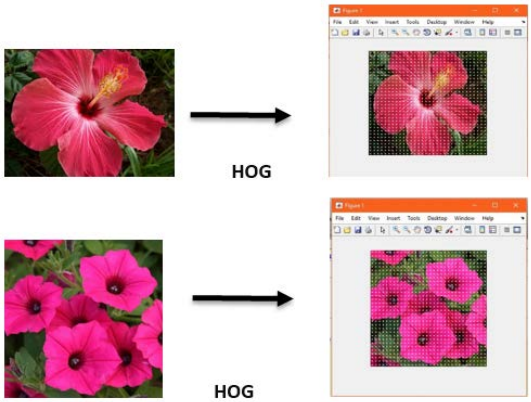
$$\text{Gradient magnitude} \quad \|\nabla F\| = \sqrt{\left(\frac{\delta F}{\delta x}\right)^2 + \left(\frac{\delta F}{\delta y}\right)^2} \tag{10}$$

$$\text{Gradient direction} \quad \theta = \tan^{-1}\left[\frac{\delta F}{\delta y} \Big/ \frac{\delta F}{\delta x}\right] \tag{11}$$

$$\text{Normalization} \quad N = \frac{\nabla F}{|\nabla F|} \tag{12}$$

HOG create a histogram of gradient orientations for each cell. Then HOG calculated normalization and descriptor blocks. To calculate the final feature vector HOG descriptors may be used for a machine learning algorithm.

Figure 3 : HOG representaion of flower images



2.3 Distance Function

For measuring distance between two points in a vector space, the thought of distance function has been established. In the part, we discuss about six distance function namely L2, chebyshev, cosine, correlation, city block and Minkowski. These distance functions are used to measure similarities between query image feature vector and the database image feature vectors.

(i) L2 Distance Function: The L2 distance between two points' p and q is the length of the line segment between them.

If $p = (p_1, p_2, \dots p_m)$ and $b = (q_1, q_2, \dots q_m)$ are two points in Euclidian n-space then the distance d between p and q is obtain by

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_m - q_m)^2}$$

$$= \sqrt{\sum_{j=1}^{m} (p_j - q_j)^2} \quad (13)$$

Where p and q is presented as Euclidian vectors. They are starting from the origin of the space with their terminal points and ending at the two points.

Euclidean norm of a is, $\|p\| = \sqrt{p.p} = \sqrt{{p_1}^2 + {p_2}^2 + \dots + {p_m}^2}$ (14)

And b is, $\|q\| = \sqrt{q.q} = \sqrt{{q_1}^2 + {q_2}^2 + \dots + {q_m}^2}$ (15)

The L2 distance between p and q is

$$\|q - p\| = \sqrt{(q-p).(q-p)} \quad (16)$$

$$= \sqrt{\|q\|^2 + \|p\|^2 - 2p.q} \quad (17)$$

Chebyshev Distance Function: Chebyshev distance function observes the total magnitude of the difference between coordinates of a couple of objects. [14]

The formula is, $d_s = |x_{mk} - x_{nk}|$ (18)

Cosine Distance function: Cosine Similarity is used for finding the similarity between the query flower image feature vector and the dataset flower image feature vectors. The similarity between two nonzero vectors of an inner product space can be measure by this function. The cosine of the angle between the vectors is measured by cosine distance function. [15]

The formula is,

$$A.B = \|A\| \, \|B\| \, cos\theta \quad (19)$$

Then the similarity is measured as,

$$Similarity, \ cos\theta = \frac{A.B}{\|A\| \, \|B\|} \quad (20)$$

$$= \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2} \sqrt{\sum_{i=1}^{n} {B_i}^2}} \quad (21)$$

Where, A_i B_i are the components of A B vectors.

Correlation distance function: Correlation measures similarity rather than dissimilarity. It is identical angular separation by centering the coordinates to its mean value.Correlation distance function

computes both linear and nonlinear association between two vectors. Calculating the distance correlation between two random vectors its compare this value to the distance correlations of many shuffles of the data.

The formula is [16],

$$d_s = \frac{\sum_{i=1}^{m}(x_{ai}-\overline{x_a}).(x_{bi}-\overline{x_b})}{\left(\sum_{i=1}^{m}(x_{ai}-\overline{x_a})^2.\sum_{i=1}^{m}(x_{bi}-\overline{x_b})^2\right)^{\frac{1}{2}}} \quad (22)$$

City block distance function: The city block distance is described as if we think about two points in the xy-plane. The city block distance is instead calculate as the distance in x plus and the distance in y. This is similar to the way instead of going straight through, if we move around the buildings in a city.

The City block distance between two points, *A* and *B,* with the dimensions m is determined as [17]:

$$City\ block\ distance\ ,\ D = \sum_{i=1}^{m}|A_i - B_i| \quad ,\qquad D \geq 0 \quad (23)$$

$$D = 0, for\ identical\ points$$

$$D > 0, for\ Showing\ little\ similarity\ points$$

Minkowski Distance Function: The Minkowski distance is a distance in a normed vector space which can be measured as a simplification of both the Euclidean distance and the city block distance.

The Minkowski distance of order n between two points [18]

$$P = (p_1, p_2, \dots, p_m)\ and\ V = (q_1, q_2 \dots q_m) \in \mathbb{R}^n \quad (24)$$

$$D(P,Q) = (\sum|p_i - q_i|^n)^{\frac{1}{n}} \quad (25)$$

When, $n < 1$, the distance between (0, 0) and (1,1) is $2^{\frac{1}{n}} > 2$

3 Experimental Results and Discussion

This section firstly discusses the image dataset construction and then the experiments to evaluate the proposed method using the dataset are carried out.

3.1 Image Dataset

Downloading the flower images from Google we make a flower image dataset. The dataset contains 300 flower images of 15 species with 25 flower images of each specie. Thus, it is difficult to detach the flower images from one another, we actually arrange our flower dataset in this way so that this type of problems can arise and we try to introduce a process to solve those problems. Figure 4 represents a portion of our dataset of different flower images.

Figure 4: A portion of flower image dataset

3.2 Retrieval Results on Individual Features

3.2.1 Color Feature Based Image Retrieval

For describing color feature we use three descriptor i.e. HSV histogram, color auto correlogram and color moments. We first calculate results for every descriptor individually.

For describing color feature we use three descriptor i.e. HSV histogram, color auto correlogram and color moments. We first calculate results for every descriptor individually.

(i) For HSV histogram feature we calculate accuracy rate of retrieving similar type of images and represent them at the following table.

Table 1: Retrieval accuracy using HSV histogram for color feature

Distance Function	Retrieval Target	No. of Retrieve Image	Average accuracy (%)
L2	**20**	**8**	**40**
City block	**20**	**8**	**40**
Minkowski	20	5	20
Chebychev	20	5	25
Cosine	20	6	30
Correlation	20	5	26

The accuracy rate of retrieving similar type of images using HSV histogram for color feature is presented by a graph at the following figure.

For color auto correlogram feature we calculate the result and represent them at the Table 2.

Table 2: Retrieval accuracy using color auto correlogram for color feature

Distance Function	Retrieval Target	No. of Retrieve Image	Average accuracy (%)
L2	20	3	15
City block	20	3	15
Minkowski	20	2	10
Chebychev	**20**	**4**	**20**
Cosine	20	2	10
Correlation	20	2	10

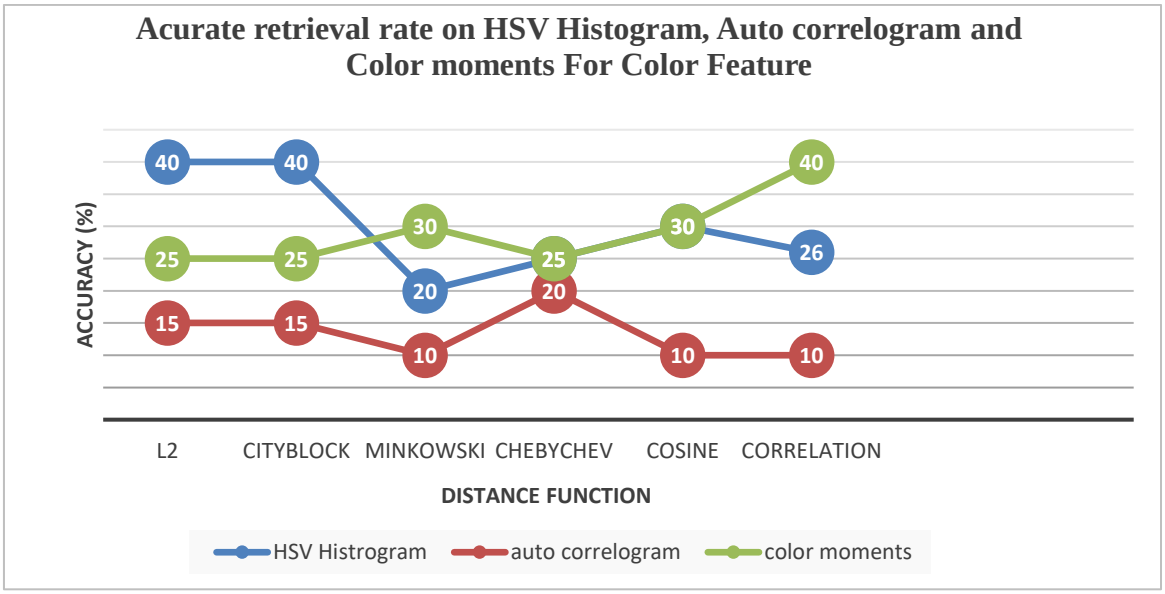
For color moments feature we calculate the result and represent them at table given bellow.

Table 3: Retrieval accuracy using color moments for color feature

Distance Function	Retrieval Target	No. of Retrieve Images	Average accuracy (%)
L2	20	5	25
City block	20	5	25
Minkowski	20	6	30
Chebychev	20	5	25
Cosine	20	6	30
Correlation	**20**	**8**	**40**

The accurate rate of retrieving similar type of images using HSV histogram, color moments and auto correlogram for color feature is presented by a graph at Figure 5.

Figure 5: Comparison of retrieval rates for HSV histogram, color auto correlogram and color moments and some distance function



Observing the above graph we can agree that the HSV histogram is better than auto correlogram and color moments. It is observable that, HSV histogram giving 40% accurate result for the distance function L2 and city block. Color moments is also a good feature and its gives 40% accuracy for correlation distance function. So it is noticeable that for the L2 and city block distance function HSV histogram feature retrieve 40% accurate images where color moments feature give the same result for the correlation distance function. The results for auto correlogram feature are poor. Its maximum accuracy is only 20% for chebychev distance function.

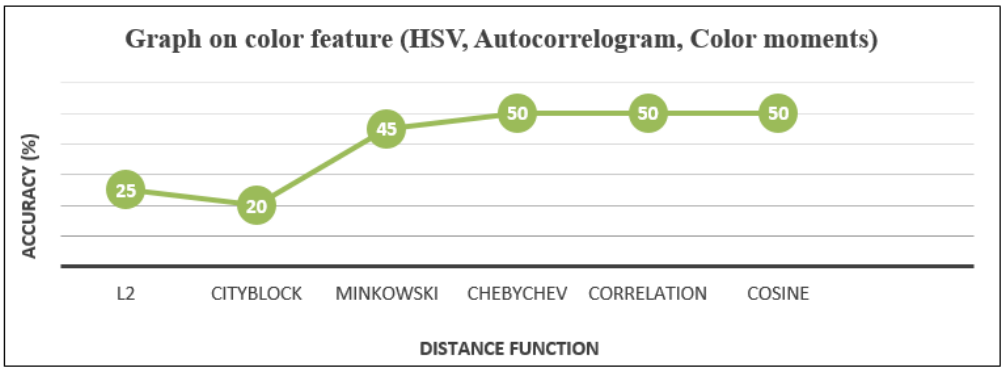
(iv) We retrieve flower images based on only their color feature. We consider the color feature as a set of three features which are HSV histogram, color auto correlogram and color Moments of the query image. Then we evaluate the results for different distance functions and present in Table 4.

Table 4: Retrieval accuracy using color feature (HSV histogram, color moments, auto correlogram)

Distance Function	Retrieval Target	No. of Retrieve Images	Average accuracy (%)
L2	20	5	25
City block	20	4	20
Minkowski	20	9	45
Chebychev	20	10	50
Cosine	20	10	50
Correlation	20	10	50

We represent the table on color feature (HSV histogram, color moments, auto correlogram) at the Figure 6.

Figure 6: Comparison retrieval rates for HSV histogram, color moments, auto correlogram and different distance function.



3.2.2 Shape Feature Based Image Retrieval

We apply our method to retrieve flower images based on only their shape feature described by HOG descriptor. Then we evaluate the results for different distance functions and construct the table 5.

Table 5: Retrieval accuracy using shape feature

Distance Function	Retrieval Target	No. of Retrieve Image	Average accuracy (%)
L2	20	3	15
City block	20	2	10
Minkowski	**20**	**4**	**20**
Chebychev	20	3	15
Cosine	20	3	15
Correlation	20	5	25

3.2.3 Combine Feature Based Image Retrieval:

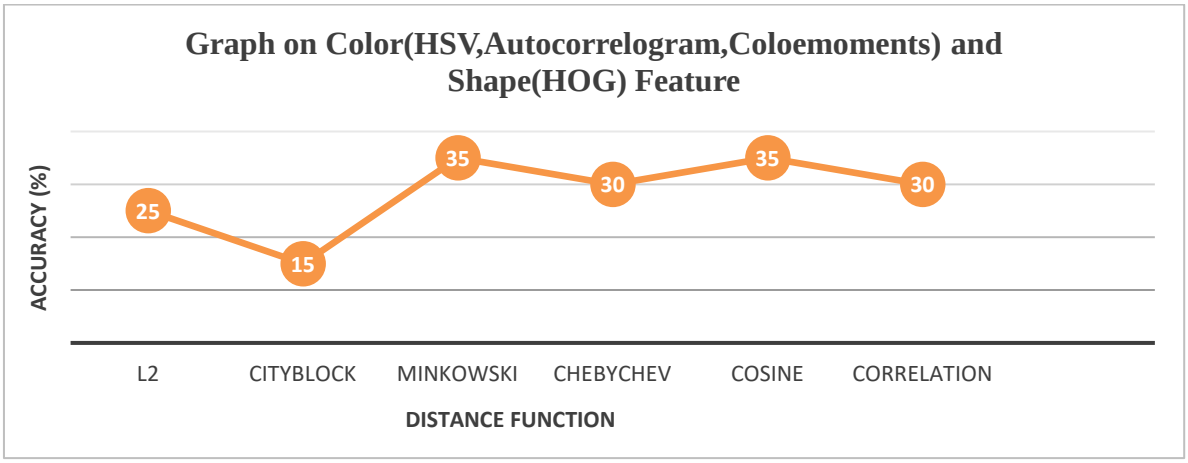
We try to retrieve flower images based on their **color and shape** features. Again we compare the results for different distance functions and construct in the table 6.

Table 6: Retrieval accuracy for combining color and shape feature

Distance Function	Retrieval Target	No. of Retrieve Images	Average accuracy (%)
L2	20	5	25
City block	20	3	15
Minkowski	20	7	35
Chebychev	20	6	30
Cosine	**20**	**7**	**35**
Correlation	20	6	30

Figure 7 is representing the accuracy rate of retrieving similar images using color and shape Features.

Figure 7: Evaluation of retrieval rates for using color (HSV, color auto correlogram, color moments), shape (HOG) features and several distance function



From the graph it is noticeable that the combination of color and Shape feature gives a good result for the correlation distance function.

4 Conclusion

In this paper, we have calculated separately the retrieval accuracy for retrieving similar type of flower images for three descriptors such as HSV histogram, color moments and auto correlogram. We observe that the result for HSV histogram is better than the other color features. We have also calculated retrieval outcomes for combining HSV histogram, color moments and auto correlogram. The retrieval results achieve an improvement for the combination of those features and better than the situation when the features are used alone. Furthermore, there is a difference in results for the different distance functions with the features. Some distance functions are giving higher results for some features only. The Minkowski and the correlation function is good for shape feature. The Minkowski and the cosine function is good when combine the color and shape features.

REFERENCES

[1] Eakins, John P., and Margaret E. Graham. "Content based image retrieval: A report to the JISC technology applications programme." (1999).

[2] Afifi, Ahmed J., and Wesam M. Ashour. "Image retrieval based on content using color feature." International Scholarly Research Notices 2012 (2012).

[3] Nilsback, Maria-Elena, and Andrew Zisserman. "Automated flower classification over a large number of classes." Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. IEEE, 2008.

[4] Varma, Manik, and Debajyoti Ray. "Learning the discriminative power-invariance trade-off." Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007.

[5] Li, Lin, and Yu Qiao. "Flower image retrieval with category attributes." Information Science and Technology (ICIST), 2014 4th IEEE International Conference on. IEEE, 2014.

[6] Kumara, YH Sharath, and D. S. Gurub. "Retrieval of flower based on sketches." Procedia Computer Science 46 (2015): 1577-84.

[7] Long, Fuhui, Hongjiang Zhang, and David Dagan Feng. "Fundamentals of content-based image retrieval." Multimedia Information Retrieval and Management. Springer, Berlin, Heidelberg, 2003. 1-26.

[8] https://stackoverflow.com/questions/4063965/how-can-i-convert-an-rgb-image-to-grayscale-but-keep-one-color.

[9] Stricker, Markus Andreas, and Markus Orengo. "Similarity of color images." Storage and Retrieval for Image and Video Databases III. Vol. 2420. International Society for Optics and Photonics, 1995.

[10] Hazra, Dipankar. "Retrieval of color image using color correlogram and wavelet filters." Proc. of International Conference on Advances in Computer Engineering. 2011.

[11] Singla, Amit, and Meenakshi Garg. "Cbir approach based on combined hsv, auto correlogram, color moments and gabor wavelet." International Journal of Engineering and Computer Science 3.10 (2014).

[12] Yang, Mingqiang, Kidiyo Kpalma, and Joseph Ronsin. "A survey of shape feature extraction techniques." (2008): 43-90.

[13] https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients

[14] https://lyfat.wordpress.com/2012/05/22/euclidean-vs-chebyshev-vs-manhattan-distance/

[15] https://en.wikipedia.org/wiki/Cosine_similarity

[16] http://people.revoledu.com/kardi/tutorial/Similarity/Correlation.html

[17] https://docs.tibco.com/pub/spotfire/7.0.0/doc/html/hc/hc_city_block_distance.htm

[18] https://en.wikipedia.org/wiki/Minkowski_distance

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 6, ISSUE 6
ISSN 2054-7412

Object Detection by Point Feature Matching using Matlab

[1]Faishal Badsha, [2]Rafiqul Islam, [3,*]Mohammad Farhad Bulbul
[1]Department of Mathematics and Statistics, Bangladesh University of Business and Technology, (BUBT), Dhaka, Bangladesh
[2]Mathematics Discipline, Khulna University (KU), Khulna, Bangladesh
[3]Department of Mathematics, Jessore University of Science and Technology (JUST), Jessore, Bangladesh
faishalbadsha0703@gmail.com; mrislam66@yahoo.com; farhad@just.edu.bd

ABSTRACT

Objects detection is an important part in image processing field. For detection object at first we take a photo called test image and then detect some region or point on the image and compare it with reference image. For this task, we use an algorithm that is used in many computer vision applications and is also considered very fast by compared to others. This algorithm can detect and describe local features for any interest object and extract features or descriptor points from it and compare it with the features that extracted from original image. Matching process is done among features and decision is made based on similar features found. This algorithm is called Speeded up Robust Features (SURF) algorithm. In this paper, we use the SURF algorithm which can detect the position of the interest object in original image by using geometric transform. This object capturing method works best for objects that exhibit in a cluttered texture patterns. When a part of object is occluded by other objects in the scene, in this case we use this algorithm to find reference image.

Keywords: Object capture, Point Feature Matching, Detection, SURF" algorithm, Cluttered Scenes, matching technique, geometric primitives.

1 Introduction

In image processing, point feature detection is an effective method to detect a specified target in a cluttered scene. Regarding specified, this method is to detect one specific object instead of that kind of objects. For instance, by using this method, we are able to recognize one specific object in a cluttered scene. The algorithm of this method is, simply speaking, based on comparing and analyzing point correspondences between the reference target image and the cluttered scene image. Object detection systems are increasingly used in the fields of industrial automation and home robotics. Real-time object learning and detection are important and challenging tasks. Among the application fields that drive development in this area, robotics especially has a strong need for computationally efficient approaches, as autonomous systems continuously have to adapt to a changing and unknown environment, and to learn and recognize new objects. For such time-critical applications, point feature matching is an attractive solution because new objects can be easily learned online, in contrast to statistical-learning techniques that require many training samples. Our approach is related to recent and efficient matching methods and more particularly to, which consider only images and their gradients to detect objects. The method of object detection works best for objects that exhibit non-repeating texture patterns, which give rise to unique feature matches. This technique is not likely to work well for uniformly colored objects, or for objects containing repeating patterns. For detecting

DOI: 10.14738/aivp.66.5619
Publication Date: 23rd December 2018
URL: http://dx.doi.org/10.14738/aivp.66.5619

objects of a particular category we use here Speeded Up Robust Features "SURF" algorithm. There is another way to detect object called "SIFT" (Scale Invariant Feature Transform) algorithm .But in our study we use here Speeded Up Robust Features "SURF" algorithm because "SURF" is more better than " SIFT ". Speeded Up Robust Features "SURF" algorithm is a local feature and descriptor algorithm that can be used in many application such as object recognition , “SURF” use much larger number of features descriptor from origin image which can reduce contribution of the errors caused by local variation in the average of all feature matching . SURF can robustly recognize and identify objects in origin images even in case of clutter and partial occlusion because "SURF" has feature descriptor which is invariant to scale, partial variant in illumination changes and orientation. The process of Speeded UP Robust Features "SURF" algorithm can be divided into three main steps. First step is "Detection step", in this step interest points are selected at distinctive locations in the origin image, such as corners, blobs etc and this process must be robustly. Second step is "Description step", in this step interest points should have unique identifiers does not depend on features scale and rotations which are called descriptor, the information of interest points represented by descriptor which are vectors that contain information about the points itself and the surroundings. Third step is "Matching step", in this step descriptor vectors are compared between the object image and the new input or origin image, the matching score is calculated based on the distance between vectors e.g. Euclidian distance and the sign of Laplacian. This algorithm can be used in many application such as recognize and locate of objects, track objects, face recognition, make 3D scenes. In this paper we used enhanced SURF algorithm for detect and locate the objects if it's found in origin image.

2 Comparison of SIFT and SURF

2.1 SIFT algorithm overview

SIFT (Scale Invariant Feature Transform) algorithm proposed by Lowe in 2004 to solve the image rotation, scaling, and affine deformation, viewpoint change, noise, illumination changes, also has strong robustness. The SIFT algorithm has four main steps: (1) Scale Space Extrema Detection, (2) Key point Localization, (3) Orientation Assignment and (4) Description Generation. The first stage is to identify location and scales of key points using scale space extrema in the DoG (Difference-of Gaussian) functions with different values of σ, the DoG function is convolved of image in scale space separated by a constant factor k as in the following equation.

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma) \times I(x, y)\big) \quad (1)$$

Where, G is the Gaussian function and I is the image. Now the Gaussian images are subtracted to produce a DoG, after that the Gaussian image subsample by factor 2 and produce DoG for sampled image. A pixel compared of 3×3 neighborhood to detect the local maxima and minima of D(x, y, σ). In the key point localization step, key point candidates are localized and refined by eliminating the key points where they rejected the low contrast points. In the orientation assignment step, the orientation of key point is obtained based on local image gradient. In description generation stage is to compute the local image descriptor for each key point based on image gradient magnitude and orientation at each image sample point in a region centered at key point; these samples building 3D histogram of gradient location and orientation; with 4×4 array location grid and 8 orientation bins in each sample. That is 128-element dimension of key point descriptor.