Automated Image Colorization with Computation of Image Colourfulness

Dr. G. Rama Subba Reddy
Associate Professor, Dept. of CSE, Mother Theresa Institute of Engineering and Technology, Palamaner, India

Abstract

Colorization is a computer based process of adding colours to a monochrome image or movie. This procedure typically includes segmentation of images in to blocks or regions and tracking these blocks across the entire image. In practice, this well-organized colorization techniques requires impressive human involvement, tedious and costly. The advancement of this process is to be fully automated colorization techniques still poses many challenges in research perspective. In this paper, we presented two techniques which are essential in the colorization process. Firstly, we are computing the colourfulness of an image useful in many image compression, camera sensor modules and to get the aesthetic quality of the images. With the help of image montage tool, we sorted out the colourfulness of the images from extremely colourful to least colourful. Along with, we also presented the scales and measures to identify the quality of image colourfulness. Secondly, unlike the conventional strategies (segmentation or region tracking) we used a deep learning strategy utilizing a pre-trained CNN to colorize the monochrome or black and white images. This model is pre-trained caffe model trained on over a millions of images and tested with the sample images. The conversion of RGB images to L*A*B colour space giving the better results compared with the state-of-art methods.

Keywords: image colourfulness, evaluation metrics, automatic colorization, pre-trained CNN.

1. Introduction

Colorization is a computerized procedure where we can add color to an image or to a movie. This procedure typically includes with the segmentation of images into different regions and also tracks its regions over the image order. Neither of such tasks is practically performed in a reliable way; consequently this colorization process requires valid user involvement and remains as a time taking and cost effective task.The complexity by colorization is however exists in a fact as it is the process which is cost effective and time taking one [1]. For instance, if an artist wants to colorize a static image initially he segments an image into some regions and later continues applying color to every region in that image. But the automatic paradigms of segmentation will often fails in finding exact fuzzy or region boundaries that are complex like a boundary between person's face and hair as well. So the artist sometimes leaves the work of representing complex boundaries between the regions. In the case of movie colorization additionally it requires finding the regions of frames of particular shot. The tracking related algorithms which are available are failed in its robustness during tracking of complex regions also needs huge user interference in process [2].

In the context of computer vision and also in machine learning aspects, the automated colorization of black and white images has been a concept for research. In this, we choose an approach known as statistical-learning-driven to clear this issue. We generate and build a Convolutional Neural Network which takes input a black and white image and produces a colorized image as an output. The system solely yields its outcome on images by having "learned from" in past in the absence of human. In the past few years, CNNs have come into action as defacto standard to solve the problems during classifying images, obtaining error rates less than 4 percent in an ImageNet challenge. CNNs achieve its success for its ability in learning and examine the colors, structures and outlines inside the images and relates those with the object classes. We hope that such features generally lend themselves in order to color the images as the object classes, structures and outlines usually correlates with the choices of colors [3, 4].

The rest of the paper is organized as follows: section 2, describes about the related works and section 3 presents the metrics used to measure the colourfulness and section 4 has a brief description about the automated colorization, section 5 describes about our experimental works and conclusions are drawn in section 6.

2. Related Work

Markel et.al, painted a color mask manually at least to get a single reference frame from a shot. Then there applied finding and tracking of movements [5], enabling colors to allocate to the other frames instantly among the regions where there will be no occurrence of motion. Color in the regions of moving are allocated with the help of optical flow, which usually needs manual fixation by operator. Even though not everybody aware about the approaches applied on industrial based colorization systems, there are some indications [6] on which such system depends on determining the regions and extracting them between shot frames. BlackMagic, which is commercial software in colorizing static images [7], enables users with purposeful tools, but the entire process of segmentation is the work of that user. In [8], the authors presented a technique to color a Grayscale image through transfer of a colour from a color of instance image. They evaluated the values of luminance of every neighbouring pixel of specific image. This approach works better on the images in which there are the regions determined with multiple varied colors results in different set of luminance or yields different textures.

Most of the Colorization paradigms are not the same in the manner they retrieve and consider the information so that to design and correspond between the Grayscale and color. When a Grayscale image is given as an input to the non-parametric approaches initially it defines one or many color version images (either user provided or instantly obtained) so as to use them as a source data. Later, follows the framework of Image Analogies, color is transformed into an input image from corresponding images of referred image(s) , Whereas parametric approaches examine prediction functions from huge color images datasets during training by rising an issue on regression onto serial color space or categorization of quantized color values. Besides our approach attempts how to perform colors classification but with a big model, huge data which is trained and also many innovations in loss function and maps to last serial output.

Our work relates to one of the parts by Ryan Dahl's CNN related system to instant image coloring. This system depends on many ImageNet-trained layers from the VGG16, combining those with the help of an instant encoder as system included with residual links which combine the intermediate outputs yielded by encoding part of network. Those residual links associated with those links prevailing in ResNet system developed by He et al won 2015 ImageNet challenge. On the basis of outcomes, this Dahl's system works in extraordinary manner in coloring the images and other objects as well. However we observed in many cases that the images are developed by system is already sepia-toned and colour muted [9, 10].

We noticed that the Dahl evaluates the colorization of an image as regression problem in which the aim of training is minimizing the total of the Euclidean distances between every pixel comprising the values of blurred color channel of image we targeted. To get better understand reason, we chose a pixel belonging to petal of a flower over many identical images. Based on this picture, that pixel can be considered with different tones like red, green, blue and so on. By the generalization of this, we hypothesize that a regression-based system would involve in developing non-saturated images and not pure with the color tone, specifically it takes more colors for the objects practically which will depict the lack of color exactness in the targeted images colorized under Dahl's system.[17, 18]

3. Image colourfulness

In recent years, Computer based image systems focus on yielding the best view of a picture instead obtaining the luminance and the colour fidelity. During the quality evaluation of processed image, we must find the difference between the actual image and the resultant image. If the quality varies it means that the quality is not good. At the time of modelling a metric for colour quality, we need to consider two key factors one is colour cast and the other is colourfulness. Here we only considered the total colourfulness of a specific image but no fidelity measurement is considered. As our work we have taken huge framework to evaluate the quality of stream of images after transmitting in network without applying any reference quality metric approach. It is only applicable on one image otherwise on a video stream but no actual image is taken [11]. In contrast we are unable to mention the quality regarding compression and coding strategy by performing an image related difference between original image and compressed image. In practise, the scheme uses Meta data which comes along with the data, for instance a parameters set that defines original image properties. Besides, the thought of not utilizing an original image to access the quality allows an approach in order to work with the images which are already underwent with tone mapping or with the algorithms of image enhancement. The degradation of colour can be done in the two ways [13]. One is through colour casts otherwise by loss of its colourfulness correlated with the colour markings based on the viewing and surrounding situations. However it is not mandatory to evaluate the complete colourfulness of specific scenery but few of the current novel techniques attempts to determine the quality of image colour in more generic framework. In order to find the answer for the question on image colourfulness we arranged psychophysical experiment, in which the subject is asked to give the rating for the colourfulness by selecting one in the 7 values. Eventually, we gave an attempt in finding an optimal algorithm which yields best output for psychophysical experiment [12].

In [15], the authors initially asked for 20 participants who are non-experts for rating the images based on 1-7 scale values of colourfulness. The scale values are not colourful, slightly, moderately, averagely, quite, highly, extremely colourful. For setting up a baseline, the researchers provided some example images along with its relative colourfulness to the participants. By the sequence of practical evaluations, they derived a common metric which correlates with the viewer's results.

Opponent color space model

Colour perception is generally not effectively represented in the case of RGB. Opponent colour model is considered as a better model of HVS. This model comprises of three components.

i. O1 is taken as luminance component
ii. O2 is a red-green channel. O2 = G – R

iii. O3 is a blue-yellow channel. O3 = B − Y = B − (R + G)

Computing the scale value

The usage of scale values enables to assume that the perceptual distance between 'slightly colourful' and 'moderately colourful' varies with the distance between 'highly colourful' and 'extremely colourful'. Since we need to assign numbers to those attributes we try with perceptual uniform scale. For instance if there is a big confusion in deciding whether it is 'slightly colourful' and 'moderately colourful', that means many images are given rating on two different categories by many people so there will be no confusion on 'highly colourful' and 'extremely colourful', it means that distance between 'highly colourful' and 'extremely colourful' is greater than the distance between 'slightly colourful' and 'moderately colourful'. We consider that the association between the types and also typical classification of that types. We initiate by forming frequency matrix in which the elements $\{K_{jg}\}$ are the how many times does an image j is placed under g category. We determine the cumulative proportion matrix comprising entries Pjg is given equation (1).

$$P_{jg}=\frac{\sum_{k=1}^{g} K_{jk}}{\sum_{k=1}^{m} K_{jk}} \qquad (1)$$

Here m denotes how many categories exists i.e., m=7. From the probability P_{jg} we derived z-scoresZ_{jg}. The relation between P_{jg}andZ_{jg} is represented as given in equation (2).

$$P_{jg}=\frac{1}{\sqrt{2\pi}}\int_{-z_{jg}}^{\infty} e^{-\frac{1}{2}\omega^2} d\omega \qquad (2)$$

Consider $t_g$as a value of remote boundary between the categories and S_j be the scale value which is not known for every category. The basic assumption lies under scale computation is shown in equation (3).

$$t_g - S_j = Z_{jg} (3)$$

The above can be represented as a matrix form and is:

$$z= X.y \quad (4)$$

$$y = [t_{1\dots}\ t_{m-1}\ s_{1\dots}\ s_m]^T \qquad (5)$$

Here z is taken as column vector comprises entire z-scoresZ_{jg}. X is the matrix which is taken to make equation (4) as equivalent to equation (3) and y is unknown value. If y is known we can know the values of the scale along with its boundaries. The values of the scale s denote the category to category distance and so the arbitrary exact value can be known. To get a result for equation 2 we suggest an extra constraint such as:

$$\sum_j s_j = 0 \qquad (6)$$

The entire scale values evaluation believes on a fact that there is some chance of getting confusion among the observers. If there are having images which get ratings at a time and they no need to give any information about scale and hence must be ignored from not being included in calculation. Eventually, we obtain scale values by solving equation 2 and then:

$$y= (X^T X)^{-1}. X^T z. \qquad (7)$$

By these calculations they observed that a basic opponent colour space representation including the mean and also standard deviations of such values are correlates approximately 95.3% of the surveyed information. Now we derive its image colourfulness metric as:

$$rg= R - G \qquad (8)$$

$$yb = \frac{1}{2}(R+G) - B \qquad (9)$$

Both of the above equations determines an opponent colour space representation where R, G and B are Red, Green and Blue respectively. In the immediate equation, yb represents half of the sum of the Red and Green channels minus the Blue channel.

Later, standard deviation (σ_{rgyb}) and mean (μ_{rgyb}) are evaluated before the final colourfulness metric calculation, C.

$$\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} \qquad (10)$$

$$\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2} \ (11)$$

$$C=\sigma_{rgyb} + 0.3 * \mu_{rgyb} \quad (12)$$

4. Automatic Colorization

In the past few years, convolutional neural networks are placed a prominent role in the area of computer vision. Here, we have three kinds of images. One is the input image and the second one is the predicted colorized image with the novel algorithm or network architecture. The third image is the original images captured. We need to compare the predicted image with the original image with the help of some colourfulness measures which are discussed in the previous section of this paper. We can consider these true images as the validation images, which never seen by the model prior to the training process. There exists some pre-trained models to extract the features from the given images to colorization process. For example, consider VGG Net architecture has simple layers, but still poses many challenges [21]. The authors in [14] proposed hyper columns in convolutional neural networks. For a specific pixel, hyper column is a vector that all the activations are above the corresponding pixel. The input image is passed through the VGG Net and extract the layers followed by upscaling and finally concatenate them together and the corresponding model is shown in Figure 1.

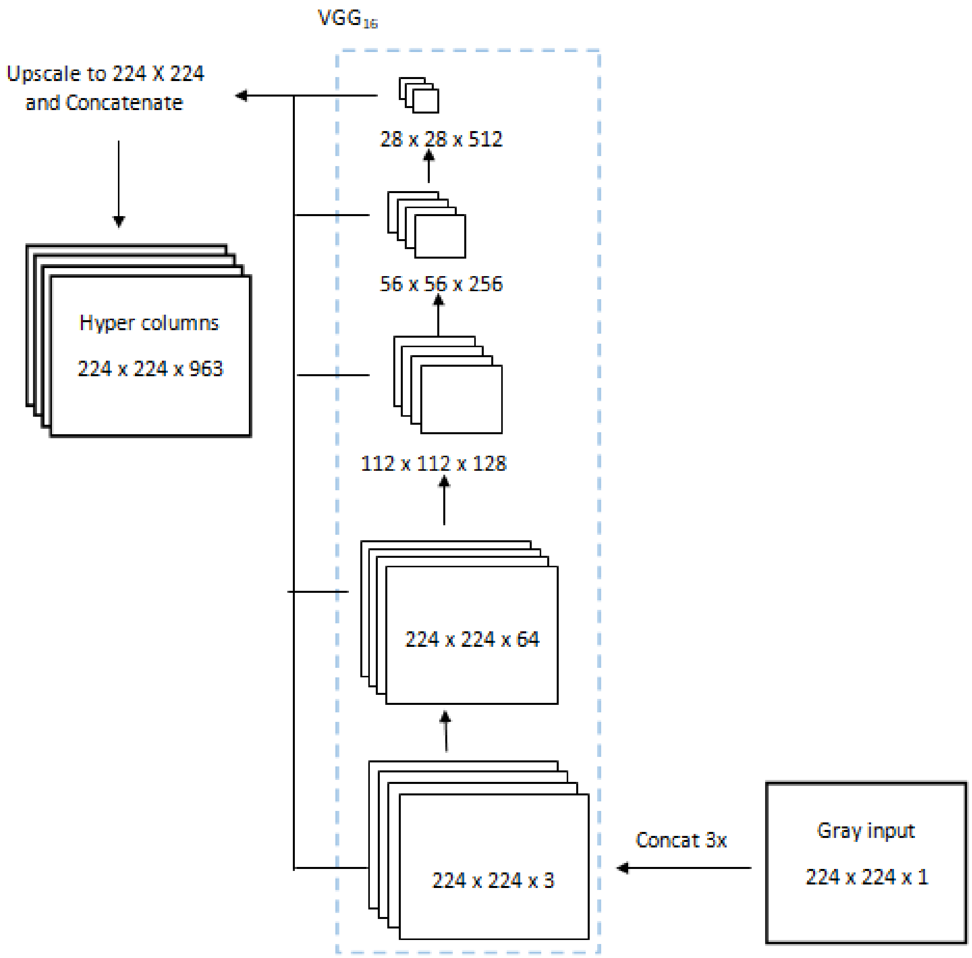

Figure 1: Convolutional layers of the model

Instead of reconstructing all the RGB images, we can train a model to produce two colour channels of different colour space such as YUV images, HSV images etc. In this paper, we also shown the results by converting the images in to LAB colour space. The YUV model is just a simple mathematical model with some basic matrix operations and easier to do the computations than the HSV colour space. The architecture used by the authors in [13] is shown in Figure 2.

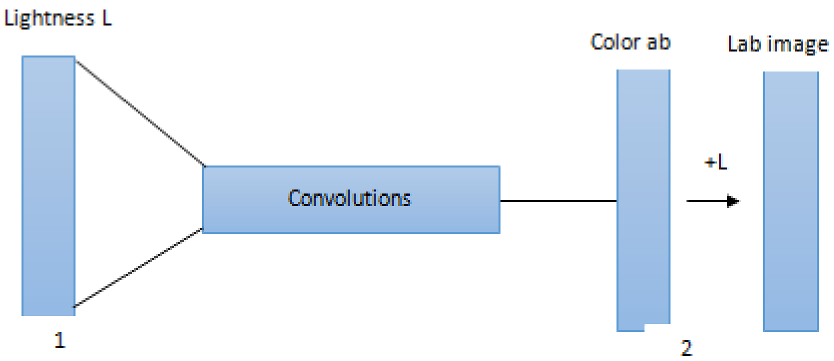

Figure 2: Architecture of Neural Network

5. Results and Discussion

There are many practical advantages by identifying the image colorfulness metric, few include in compression algorithms, vision sensors, to compute aesthetic quality of images and in many other image visualization techniques. The scale values are not colorful, slightly colorful, moderately colorful, averagely colorful, quite colorful, highly colorful and extremely colorful. These values are ranging from 1 to 7. There are numerous metrics to correlate the results of the viewers for example computing the mean and standard deviation etc. we implemented this work using OpenCV and numpy. As a first step, split the original image in to three color channels and Compute the opponent image colorfulness metrics '**rg**' and '**yb**' as per equation (8) and (9) followed by the computation of mean and standard deviation. [Here 'rg' denotes red-green channel, 'yb' denotes the yellow and blue channel]. Then combine the mean and standard deviations of 'rg' and 'yb' colorfulness metrics. Final computation metric is evaluated as shown in equation (12). The images are resized to image the pixel size of 250 by maintaining the aspect ratio. We also sorted the result set in the form of most colorful to the least colorful. In our experiment, we have taken the UK bench data set, the most popular dataset which contains nearly 11000 images of around 2550 groups of images with size 640 X 480. In which we have taken 2000 sample images. For these images we have computed the colorfulness metric from the least colorful to most colorful. The metrics for few images (i.e. 10) are shown in Figure 3. In other words, we stored the top 10 of least colorfulness images and most colorfulness images as shown in Figure 3.a and Figure 3.b.

(a)

(b)

Figure 3: Image colorfulness a. least colorful b. Most colorful.

On the other hand, when the colorfulness of the image is too low (i.e. black and white images), there is a requirement to colorize the images. With the help of deep learning algorithms, it is easy to train the images and colorized it with the help of some convolutional layers. In simple terms, image colorization is the process of converting the gray level images in to color. The semantic colors of the images may be vary according to the environmental conditions. For example, ocean color on a heavy sunny day is “blue”. The existing methods involves manual or human interaction and in some cases it leads to desaturated colors. The novel approach is to use the convolutional neural networks with the help of deep learning we can automatically colorize the new images with better accuracy. We performed the experiments by taking the sample images and applied the deep learning mechanism and have shown sample results in Figure 4. The part of the work is taken from [2], in which the colorful image colorization was done with the help of colorization algorithm. The work has taken the sample images from ImageNet and are converted to L*A*B color space. It is similar to RGB model in which the L denotes the lightness channel, ‘a’ paired up with green-red and ‘b’ paired up with blue-yellow. The channel ‘L’ may be taken as the gray scale image and it is considered as the input to the architecture. As a second step, we can give all these images to the network and train the network to get the prediction results. Later, the ‘L’ channel can be merged with the other ‘ab’ channels and this L*A*B image

a b

Figure 4: Automated colorized results (left : input , right: output)

Converted back to the original RGB image. In Figure 4, the left hand side images are input and the right hand sides images are the output images which are automatically colorized by the model. The images are collected from the internet and the images are never seen by the model before training.

6. Conclusion

Computer based colorization is the process of adding colors to the monochrome images or video frames. The two major contributions of this paper include the techniques which are useful to colorize the images automatically without interfering the humans in to the process. At first, the colorfulness of the image is computed broadly in to two categories; i.e. least colorized and most colorized. This technique is mainly based on the computation of mean and standard deviations of the opponent color space. We presented the metrics to compute the colorfulness of an image. With the help of this method one can easily find the colorfulness in a less computational time. Later, we studied the novel methods introduced in the recent years along with neural network architecture. Then, with the help of CNN using deep learning monochrome images are colorized automatically by giving the test images to the pre-trained model. The images were collected randomly from the internet, and never seen by the model before the training process. The method, not only applicable for the monochrome images, but can also applied to the black and white video streams.

References

[1] Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. European Conference on Computer Vision (2016)

[2]. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 415–423

[3]. Jayaraman, D., Grauman, and K.: Learning image representations tied to ego-motion. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1413–1421.

[4]. Lotter, W., Kreiman, G., Cox, D.: Deep predictive coding networks for video prediction and unsupervised learning. arXiv preprint arXiv:1605.08104 (2016)
[5] MARKLE, W., AND HUNT, B., 1987. Coloring a black and white signal using motion detection. Canadian patent no. 1291260, Dec.
[6] SILBERG, J., 1998. The Pleasantville post production team that focussed on the absence of color. Cinesite Press Article, http://www.cinesite.com/core/press/articles/1998/10 00 98- team.html.
[7]. WELSH, T., ASHIKHMIN, M., AND MUELLER, K. 2002. Transferring color to greyscale images. ACM Transactions on Graphics 21, 3 (July), 277–280.
[8]. NEURALTEK, 2003. BlackMagic photo colorization software, version 2.8. http://www.timebrush.com/blackmagic
[9]. Dahl, R.: Automatic colorization. In: http://tinyclouds.org/colorize/. (2016)
[10]. Deshpande, A., Rock, J., Forsyth, D.: Learning large-scale automatic image colorization. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 567–575
[11]. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, and A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915 (2016)
[12] Nister, David, and Henrik Stewenius. "Scalable recognition with a vocabulary tree." Computer vision and pattern recognition, 2006 IEEE computer society conference on. Vol. 2. Ieee, 2006.
[13]. Zhang, Richard et al. “Colorful Image Colorization.” *ECCV* (2016).
[14]. Hariharan, Bharath et al. “Hypercolumns for object segmentation and fine-grained localization.” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2015): 447-456.
[15]. Hasler, David and Sabine Süsstrunk. “Measuring colourfulness in natural images.” (2003).
[16]. Aditya Deshpande, Jason Rock and David Forsyth. Learning Large-Scale Automatic Image Colorization. In *ICCV*, Dec 2015.
[17]. Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep Colorization. In *ICCV*, Dec 2015.
[18].Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning Representations for Automatic Colorization. In *ECCV* 2016.
[19]. Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. In *SIGGRAPH*, 2016.
[20]. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. ACM Transactions on Graphics (Proc. of SIGGRAPH 2016) 35(4) (2016)
[21]. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)

Author Biography:

Dr. G. RamaSubba Reddy received his Ph.D in Computer science & Engineering from Sunrise University. He received his M.E in Computer Science & Engineering from Sathyabama University, Chennai. Presently he is working as an Associate Professor and Head of the Department in Computer Science & Engineering, Mother Theresa Institute of Engineering & Technology. Palamaner, Andhra Pradesh. His current research focus is on IoT, wireless sensor Network, Data mining, and Cloud computing.