

Film Colorization, Using Artificial Neural Networks and Laws Filters

Mohammad Reza Lavvafi

Department Computer, Islamic Azad University of Mahallat
Mahallat, Arak, Iran

S. Amirhassan Monadjemi and Payman Moallem

Department of Computer Engineering, Department of Electrical Engineering
Faculty of Engineering, University of Isfahan
Email: monadjemi@eng.ui.ac.ir

Abstract—In this study a new artificial neural network based approach to automatic or semi-automatic colorization of black and white film footages is introduced. Different features of black and white images are tried as the input of a MLP neural network which has been trained to colorize the movie using its first frame as the ground truth. Amongst the features tried, e.g. position, relaxed position, luminance, and so on, we are most interested on the texture features namely the Laws filter responses, and what their performance would be in the process of colorization. Also, the network parameter optimization, the effects of color reduction, and relaxed x-y position of pixels as the feature, are investigated in this study. The results are promising and show that the combination of MLP and texture features is effective in this application.

Index Terms—Colorization, Artificial neural networks, Laws filters, Color reduction, Texture features.

I. INTRODUCTION

Colorization or adding color to the black and white images and movies, has widely spread since 1980 and had many adherents in the film industry and computer graphics. Colorization is more important in some fields such as colorization of monumental images, medical diagnosis (e.g. pseudo colored tomography images, or MRI), colorization of black and white classic or documentary movies, historical reprographics, and so on [1][6].

In all of the colorization processes, the main idea is replacing saved gray level of each pixel of black and white image (i.e. luminance) with a color vector in the cubical color space, for example, red, green, and blue cubical vector in the RGB (Red-Green-Blue) color space which requires a sort of intelligence. So, in most colorization processes, we will need human interference at least in a part of the process to determine the correct color.

In this research, we have presented a black and white video footages colorization method using artificial neural networks along with the digital image processing techniques that have tried to reduce the interference of the

human operator in the colorization. While the ultimate goal is getting through a fully automatic, high precision, and robust colorization method of frame images.

Wilson Markel was the first one who did colorize the black and white (B/W) movies in the early 80's. Since then, several manual or semi-automatic colorization methods have been developed for black and white movies and image colorization. In most of these methods, a black and white image is divided into some parts and then, color is transferred to it through the source of color image, or human user will specify color of these determined areas and sections, and the colors spread through the whole area next [1][3].

Colorization process is relatively time consuming and costly. Although these two major problems are disappearing little by little by advances of both hardware and algorithms, they are still main problems facing the process. In the proposed scheme, we are going to introduce an automatic process to colorize the black and white movies using ANN. Although much time is needed to train ANN, it is possible to decrease the consuming time by employing faster processors or ANN training algorithms. Also, if with each colorized frame and trained neural networks we can colorize a sequence of 50 frames in average, that is nearly 2-3 seconds of a movie, nevertheless; this colorization process is approximately 50 times faster compared to those methods where all frames should be colorized manually.

In the performed scheme, we firstly colorize a source B/W frame manually [4]. Then, a Multi Layer Perceptron (MLP) neural network will be tried using these 2 frames. (B/W frame as the input, and the original color frame as the desired output). Afterwards, data of the sequential black and white frames in series will be placed on the input of the neural network and the corresponding color data of those frames will be appeared on its output.

As mentioned before, in this research, we used no fully automatic algorithm to colorize the source frame. However, to examine the performance of the utilized method more closely, we firstly turned the color movie into black and white, and then this black and white movie

re-colored to calculate the error of the utilized scheme. In other words, the performance of colorization method would be evaluated considering the original color movie as the reference.

This paper is organized as follows: In Section II the background of the study would be discussed. Next in Section III the new method is introduced in details. The performance of our method is evaluated in Section IV. Finally, the results are discussed in Section V and the paper is summed up in Section VI.

II. BACKGROUND

In one of the first colorization processes which developed by Markel and Hant, an image (frame) was selected from each scene of a movie and colorized manually, then colors were transferred and spread automatically to the next frames of the same scene where there were no motions. Colors in moving areas sometimes needed to set manually again by the operator [3][5].

Most colorization algorithms are based on the colorizing of regions in a frame, and following these regions in the subsequent frames. For instance, Black Magic is a commercial software introduced in 2003 for image colorization which is based on the neural technology. In this software, color tables and many patterns are provided for the human users and image partitioning is also done by the user [3][10].

The first colorized movies applying this technique were colorless, transparent, low contrast, and had washed out colors. Although this technology developed to a large extent in the 80's and colorized successfully some television broadcastings, it suffered from high cost and personnel demand, and complicated and troublesome procedures. Also they were so operator-oriented and the quality of colorization was totally proportional with the operator's skill [1][3][6].

Welsh presented a semi-automatic technique to colorize black and white images in 2002. According to Welsh, colors transferred from a source color image to black and white one (target image) and the luminance to each pixel is measured in a black and white image and color in the source image transferred to the target one from a pixel with the most similar luminance. This method was worked well on the images that had regions with various colors and different luminance. In other images, the user had to find the corresponding pixels in both source and target images directly. Although this method resulted well, the user's interference was again of essence. Also, the user needs to find the source image containing suitable colors on the regions, and then relates some parts of the target image that are to be colorized with the source image. So, getting good results in some parts of the image is difficult and required high perseverance and proficiency of the user. His technique does not certify harmony of the colors, and in some images, it assigns completely different colors to the pixels with similar brightness. Experimental results suggest that extreme similarity of the source image to the target one had special effects on the outcome [1][2].

Next, our proposed fully automatic colorization technique of black and white images will be described briefly.

III. METHOD AND TOOLS

A. The procedure

The proposed method has got two main stages:

- 1- Finding the most appropriate key color images.
- 2- Colorization of black and white frames using the selected key color image, visual features of the B/W frames, and an artificial neural network (ANN).

Basically, we attempt to train an ANN to perform the colorization. To train the ANN we need pairs of luminance-color as the input-output for each pixel of the training image. The first frame of a movie can be colorized manually and used as the key (training) image. This, however, means having a not fully automatic colorization. Instead, we can find a color key image within a predefined image database which of course is not the focus of this paper.

The main goal of this research is to develop an intelligent automatic colorization of black and white movies using artificial neural networks. After studying different types of ANNs and their applications, we concluded to use multilayer Perceptron neural networks. Since these networks are of the most generic ANNs to explain some problems like ours, that is in fact a mapping from a lower dimensional black and white to a higher dimensional color space. From now on, our purpose of neural networks in this text will be a multilayer Perceptron ANNs.

In our method, at the beginning, we manually colorize one frame of the black and white movie (source frame), and then designed neural network will be trained based on this frame. Finally, the rest of the frames will be colorized employing the trained neural network.

In this special problem, we fed our neural networks with all the pixels in the key frame without any real preference amongst the pixels. So, the employed neural networks need to be designed to take a single pixel instead of a set of frames as the input. The network has got no knowledge about the spatial relationship between pixels, unless it has been modeled by a sense in the pixel's feature vector. In other words, characteristics of a pixel in a given black and white frame considered as the input of the neural network and color components of that pixel in the colorized frame seem to be its output, and we have to colorize all the pixels of a frame orderly and one by one. In this study, to specify the performance of neural networks in movies colorization, we converted a few fragments of short color movies with some 100 to 300 frames to black and white. Then, we colorized these black and white fragments using our algorithms.

To evaluate the performance of different proposed colorization schemes and feature sets, besides the visual inspection of the colorized footages, a Mean Square Error (MSE) rate was defined and calculated for each colorized

movie. Mentioned later at Section IV, this error rate can help us to methodically evaluate our method and settings.

we followed out the whole tests in RGB color space, since it is the most popular color space exist. The neural network used was a MLP with one hidden layer, sigmoid/linear squashing function for hidden/output layer neurons, and a Scaled Conjugate Gradient (SCG) training scheme. The optimum number of neurons in the hidden layer was approximated by trial and error and set to 50.

B. Research outline

In all processes, input and output of the utilized neural networks are normalized between [0, 1]. We examined those tests on two movies with AVI format and 640×480 resolution (namely *man.avi* and *plane.avi*). Common characteristic of both movies is slowness of changes in their consecutive frames. A few frames of these two movies are shown in Figure 1 and 2.

Practically, we started with a MLP [9] [13] with 1 input, 3 outputs, and 50 neurons in the hidden layer. The unique feature was the gray level of the pixels. 3 outputs indicate the R-G-B levels of the colorized pixel. Next, we added the $\langle x, y \rangle$ Cartesian coordinates of the pixels and came across a 3-featured, $3 \times 50 \times 3$ MLP. Shown in Table 1 and discussed later, the improved results are compared to the first test. *Other types of ANN may be tried too, however, the MLP performance was promising and it did not seem necessary regarding this particular test suite.*

The basic idea to select those features is the importance of gray level and coordination as the only information directly in hand about a pixel. Also, regarding the slight changes between two serial frames, the $\langle x, y \rangle$ coordinate would be close in a small neighborhood of those frames, where we rationally expect similar colors too. This by a sense hints us to involve the texture. Although texture relates to the material of a surface, it includes the effect of neighborhood too. To extract the textural features we employed the Laws filters [11].

Introduced by the name sake in the early 80's, a classic Laws filter bank contains 9 filters that have been determined experimentally and it is believed that we can analyze the textural characteristics of an image by applying those filters. So it was tried to use the coherence principle in colorization process using these filters. Laws filter responses of pixel of each B/W frame were used as the new features for colorization in some of our tests.

Regarding the slight changes between the coordinates of any pixel in two consequent frames, we employed the 5×5 super pixels instead of normal pixels to develop a grid, lower resolution, coordinate system for the image. Thus, in some further experiments coordinate features were replaced with super pixel coordinate ones, where any pixel in a $5 \times 5 = 25$ neighborhood is assigned the same $\langle x, y \rangle$ coordinate features.

Colorization is a mapping problem indeed: from a 256 gray level space to a $2^{24} = 16M$ RGB one. Decreasing the dimension of the destination space will help the ANN to perform a more robust mapping. While, due to the

limitations of the human vision system (defined as the Just Noticeable Difference or JND [12]), much less than 256 colors would be noticeable for a human observer in each R-G-B channel. Thus, we decided to decrease the number of colors in both key/training and colorized frames from 16M to $2^{12} = 4096$ (4K). To keep it as simple and fast as possible, for each pixel we removed the 4 lighter bits in each color channel and get through some $2^4 \times 2^4 \times 2^4 = 4096$ colored frames. Having 4096 colors in the movie did not affect the observer, is balanced amongst the color channels, and is easy to implement. Further color reduction caused some obvious degradation in the footages. Generally however, compared to the original movies, color reduction in different ways may lead us to more proper outcomes.

Further attempt to improve the colorization results was carried out using three individual ANNs, one for each color channel. In this setup, three separate ANNs were trained with one of the R, G, and B levels of the pixels respectively.

Finally, different combinations of the above mentioned submethods and feature sets were tested and compared to find out the best colorization schemes. The results will be discussed later in Section V.

IV. PERFORMANCE EVALUATION

To evaluate the colorization performance mathematically and analytically, the MSE rate in each color channel and between two given frames, A and B, is computed as [7][8][12]:

$$MSE_R = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w (R_{ij}^A - R_{ij}^B)^2 \quad (1)$$

$$MSE_G = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w (G_{ij}^A - G_{ij}^B)^2 \quad (2)$$

$$MSE_B = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w (B_{ij}^A - B_{ij}^B)^2 \quad (3)$$

In those equations above, h and w indicate the horizontal and vertical frame size in pixel. We used the relation below to calculate the MSE rate of all frames of the movies:

$$MSE_R = \frac{1}{(h \times w) \times nf} \sum_{k=1}^{nf} \left[\sum_{i=1}^h \sum_{j=1}^w (R_{ij}^{f_1} - R_{ij}^{f_2})^2 \right]_k \quad (4)$$

$$MSE_G = \frac{1}{(h \times w) \times nf} \sum_{k=1}^{nf} \left[\sum_{i=1}^h \sum_{j=1}^w (G_{ij}^{f_1} - G_{ij}^{f_2})^2 \right]_k \quad (5)$$

$$MSE_B = \frac{1}{(h \times w) \times nf} \sum_{k=1}^{nf} [\sum_{i=1}^h \sum_{j=1}^w (B_{ij}^{f_1} - B_{ij}^{f_2})^2]_k \quad (6)$$

$$MSE_{total} = \frac{MSE_R + MSE_G + MSE_B}{3} \quad (7)$$

In the equations above, *nf* indicates the number of frames, *k* is the frame indicator, and *f1* and *f2* are movies to compare. Equation 8 is used to evaluate the error percentage, considering 256 different levels of each color channel:

$$MSE_{total} \% = \frac{MSE_{total}}{256} \times 100 \quad (8)$$

Finally, we applied MSE rate in (9) to evaluate the colorization performance of the utilized method:

$$Performance = \frac{MSE_{total}^{BW-C} - MSE_{total}^{C-Cz}}{MSE_{total}^{BW-C}} \times 100 \quad (9)$$

Where MSE_{total}^{BW-C} means the total rate of error between B/W and the original color movies. Likewise, MSE_{total}^{C-Cz} shows the total rate of error between the colorized and original color footages.

V. RESULTS

To avoid the ANN's over fitting problem, after each 10 epochs of training, the 2nd frame is fed to the network as the evaluation criterion. If the MSE moves high, the training will be stopped. Some other useful penalty based anti-over fitting techniques were described in [14]. All the accomplished tests and their performance have been summarized in Table1. The columns of that table show the test number, features used on that particular test, and the colorization performance (percent) for two test footages, namely Plane.avi and Man.avi (See Figure 1 and 2). In the first row, Test #1, the unique feature is the gray level of pixels, and the colorization performance is limited to 48% at most.

In Test #2 to #4, <x,y> coordinates, Laws filter responses plus <x,y> coordinates, and super pixel coordinates are added to the gray levels respectively. Although slight improvements materialize, the performance still is not promising. Next in Test #5 to #7 the first promising results are achieved after a 16M to 4K color reduction (underlined). Test #9 to #16 are in fact repetitions of the former eight tests, besides this time three separate ANNs are employed, one per color channel. The best colorization performance is obtained by Test #15, where separate ANNs are fed by gray level, <x,y> coordinate, and the Laws filter responses of pixels in a 4096 colored version. The average performance of almost 89% is achieved on both sample movies. Figure 1

and 2 illustrate the results of the 15th test on two sample footages respectively. The visual quality of the proposed colorization can be reviewed through the presented frames.

Apart from the first subset of tests (#1 to #4), super pixel features do not show a high performance compared to the Laws filters responses. This suggests that the Laws implicit but adjustable and wide neighborhood effect is more reliable than the fixed 5x5 neighborhood correlation of the super pixels in colorization applications.

VI. CONCLUSION

We can summarize the results of the presented study as follows:

The basic colorization feature of each pixel is its gray level; nevertheless, using Cartesian coordinates of pixels as another colorization feature will be undeniable and improve the performance.

TABLE I. EXPERIMENT RESULTS

Test#	Features / ANN input	Colorization performance (percent)	
		Plane.avi	Man.avi
1	Gray level of pixels	48%	24.3%
2	Test #1 + Cartesian coordinates of pixels	58.44%	51.7%
3	Test #2+ 9 Laws filter responses	54.54%	44.6%
4	Test #1 + coordinates of super pixels	67.48%	50.1%
5	Test #1, color reduction to 4096	90.03%	61.2%
6	Test #2, color reduction to 4096	<u>92.17%</u>	<u>84.7%</u>
7	Test #3, color reduction to 4096	<u>93.43%</u>	<u>81.9%</u>
8	Test #4, color reduction to 4096	32.85%	18%
9	Test #1, separate NN for each color channel	47.8%	25.9%
10	Test #2, separate NN for each color channel	66.92%	54.5%
11	Test #3, separate NN for each color channel	64.73%	51.24%
12	Test #4, separate NN for each color channel	67.8%	42.5%
13	Test #5, separate NN for each color channel	12.5%	14.7%
14	Test #6, separate NN for each color channel	<u>94.21%</u>	<u>82.8%</u>
15	Test #7, separate NN for each color channel	<u>95.71%</u>	<u>83.56%</u>
16	Test #8, separate NN for each color channel	29.7%	17.46%

Using Laws filters, as filters that make coherence principle of adjacent pixels involved in the colorization process, slightly improves the performance too. However, it needs to be considered that training and testing process of neural networks will take longer time after adding each feature. Therefore, using these filters will not be recommended when the time is of importance. In separate tests, when we examined the effects of Laws filters without considering the Cartesian coordinates of pixels, (i.e. just having gray level of pixels and Laws filters response as features), the colorization performance decreased drastically. This also confirms the effective role of Cartesian coordinates in the colorization process.

Color reduction of the original movie from 16 million to 4096 colors (i.e. having 16 colors in each RGB channel instead of 256) had a considerable and positive effect on the colorization performance. Therefore, it can be said that color reduction must be carried out as one of the necessary preprocessing steps. *Tests #1, #5, and #13 use the same single "gray level of pixel" feature in three different ways. From test #1 to #5 we can see a promising increase in the performance, (e.g. from 36.15% to 75.6% in average). However in test #13 the performance jumps very lower to 13.6% in average. The reason can be the single gray level feature, where using three separate ANNs and feeding them with the same feature just adds to the redundancy.*

Based on the results of this research, we can make the following suggestions:

- Applying other color spaces instead of RGB color spaces such as CIE-Lab, HIS or YIQ color spaces. Because in these color spaces, color components will be different from luminance component and neural network output will only be limited to 2 color components.
- Using more effective methods than Laws filters to analyze the image texture and adjacency of pixels. However, image pixels have not only spatial correlation but also time correlation in sequential frames, using more appropriate frames may lead to better results.
- It will be better to find and colorize an object in sequential frames instead of colorization of pixels or super pixels in each frame. If system will be able to recognize the image objects, it may be more successful in their colorization applying their color in previous frames.
- Study of neural networks with various structures instead of Perceptron. We selected this neural network since multilayer Perceptron neural networks would be more general than other structures of neural network in problem solving. Other types of neural network may respond more appropriately to our problem [9][13].



Figure 1. Plane.avi colorization outcomes, for four different frames, using Test #15 conditions.



Figure 2. Man.avi colorization outcomes, for four different frames, using Test #15 conditions.

REFERENCES

- [1] Welsh T., Ashikhmin M., Mueller K., "Transferring color to grayscale images", ACM Transactions on Graphics Vol.20, No.3, July 2002, pp. 277–280.
- [2] Irony R., Cohen-Or D., Lischinski D. "Colorization by Example", Proceeding Eurographics Symposium on Rendering, 2005, pp201-210.
- [3] Levin A., Lischinski D., Weiss Y. "Colorization using Optimization", ACM Transactions on Graphics Vol. 23, No. 3, 2004, pp 689–694.
- [4] Qiu G., Guan J., "Color by Linear Neighborhood Embedding", IEEE International Conference of Image Processing – Genova – Italy, september 2005, pp 988-991.
- [5] Di Blasi G., Reforgiato Recupero D. "Fast Colorization of Gray Images", Proceeding Eurographics Italian Chapter, 2003, http://svg.dmi.unict.it/iplab/administrator/users/rg_it2005.pdf.
- [6] Vieira L.F., do Nascimento E., Fernandes Jr F., Carceroni R., Vilela R., Araujo A., "Fully automatic coloring of grayscale images", Image and vision computing magazine Vol. 25, 2007, pp 50-60.
- [7] Egmont Petersena M., de Ridderb D., Handelsc H., "Image processing with neural networks—a review", Pattern Recognition magazine Vol.35, No. 10, 2002, pp 2279-2301.
- [8] Sandberg K., "Introduction to image processing in Matlab"; Department of Applied Mathematics, University of Colorado, <http://www.colorado.edu>.
- [9] Sigmon K., "MATLAB Primer"; Department of Mathematics University of Florida, 1993, <http://www.glue.umd.edu/~news/ench250/primers.htm>.
- [10] Wikipedia, "Film Colorization", Retrieved from: <http://en.wikipedia.org/wiki/colorization>, 2007.
- [11] Monadjemi A., "Towards Efficient Texture Classification and Abnormality Detection"; PhD Thesis, University of Bristol, 2004, pp 19-21.
- [12] Gonzales R., Woods R. "Digital image processing"; Prentice Hall, 2007.
- [13] Menhaj, M. B. "An introduction to neural networks"; Amirkabir University of Tech, 2000.
- [14] Woo W.L. and Dlay S.S., "Regularised nonlinear blind signal separation using sparsely connected network", *IEE Proc. on Vision, Image and Signal Processing*, vol. 152, no. 1, 2005, pp. 61-73.

Mohammad Reza Lavvafi, born 1970, in Isfahan, Iran. He got his BSc in computer engineering, hardware from Isfahan University of technology in 1992, and MSc from Islamic Azad University of Najafabad in 2007. He is now working as a lecturer at Islamic Azad University, Mahallat, Iran.

Seyed Amirhassan Monadjemi, born 1968, in Isfahan, Iran. He got his PhD in computer engineering, pattern recognition and image processing, from University of Bristol, Bristol, England, in 2004. He is now working as an assistant professor at the Department of Computer, University of Isfahan, Isfahan, Iran. His research interests include pattern recognition, image processing, and human/machine analogy.

Payman Moallem, born 1970, in Tehran, Iran. He is assistant professor in Electrical Engineering Department of University of Isfahan, Iran. He received his B.S. and M.S. both in Electronic Engineering from Isfahan University of Technology and Amirkabir University of Technology, Iran, in 1992 and 1995 respectively. He also received his PhD in Electrical Engineering from Amirkabir University of Technology in 2002. From 1994 to 2002, he has researched in Iranian Research Organization, Science and Technology (IROST) on the topics like, parallel algorithm and hardware used in image processing, DSP based systems and robot stereo vision. He has published over 26 papers in refereed journals. His interests include fast stereo vision, target tracking, real-time video processing, neural network and image recognition and analysis.