Centro de Informática
U·F·P·E
Pós-Graduação em Ciência da Computação

Saulo César Rodrigues Pereira Sobrinho

Beyond Landscapes: An Exemplar-based Image Colorization Method

Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
http://cin.ufpe.br/~posgraduacao

Recife
2018

Saulo César Rodrigues Pereira Sobrinho

Beyond Landscapes: An Exemplar-based Image Colorization Method

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Processamento de Imagens
Orientadora: Judith Kelner

Recife
2018

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

P436b Pereira Sobrinho, Saulo César Rodrigues
Beyond landscapes: an exemplar-based image colorization method / Saulo César Rodrigues Pereira Sobrinho. – 2018.
65 f.: il., fig., tab.

Orientadora: Judith Kelner.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2018.
Inclui referências.

1. Inteligência artificial. 2. Colorização de imagens. I. Kelner, Judith (orientadora). II. Título.

006.3 CDD (23. ed.) UFPE- MEI 2019-009

Saulo César Rodrigues Pereira Sobrinho

Beyond Landscapes: An Exemplar-based Image Colorization method

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 19/12/2018

BANCA EXAMINADORA

____

Prof. Dr. Carlos Alexandre Barros de Mello
Centro de Informática/UFPE

____

Prof. Dr. Alejandro César Frery Orgambide
Instituto de Computação/UFAL

____

Profa. Dra. Judith Kelner
Centro de Informática/UFPE
(**Orientadora**)

ACKNOWLEDGEMENTS

Esta dissertação marca a conclusão de mais uma etapa importante da minha formação e, no cumprimento desta etapa, gostaria de deixar agradecimentos à pessoas que tiveram particular importância nesta caminhada.

Primeiramente, gostaria de agradecer a todos os professores que fizeram parte da minha formação, desde as etapas iniciais na escola, onde me ajudaram a construir a fundação até no período mais recente na universidade, onde pude absorver muito conhecimento e experiência profissional.

Gostaria de agradecer à minha orientadora, professora Judith, não só pela contribuição na função de orientadora da pesquisa, mas também pela compreensão e solicitude para cumprir esta função com a urgência que foi necessária para que eu pudesse seguir meus planos para o próximo ano.

Também gostaria de mencionar os membros da banca, os professores Carlos e Alejandro, pela leitura atenciosa e pelas valiosas sugestões de melhorias para este trabalho.

Aos atuais e antigos colegas do grupo, agradeço pelo conhecimento compartilhado e pela boa convivência e boas conversas nos almoços ao longo dos anos.

Aos também colegas de grupo e amigos Santos, pelo apoio e camaradagem ao longo de toda a jornada universitária, e meu primo Pedro, pela amizade, pelas incontáveis caronas e por atender minha chamada repentina para ser minha testemunha.

Por último mas não menos importante, gostaria de agradecer à minha família. Ao meu irmão, por ser uma fonte constante de aprendizado para mim e por encher meu juízo para que eu terminasse logo esse documento. Aos meus pais, por terem sido os responsáveis pela minha educação e formação pessoal e por todo o suporte, não apenas durante esta etapa mas em toda minha vida. E finalmente, à minha esposa, que foi fundamental na elaboração deste trabalho, tanto pelas sugestões e críticas valiosíssimas sobre a pesquisa como pelo enorme apoio e paciência em todo o processo.

ABSTRACT

Image colorization consists in, given a grayscale image, generating a plausible color version of this image, which can be performed as a manual/artistic process[1] but also as a computer assisted or even fully automated process. Colorization is a underconstrained problem, which requires extra information in order to provide a unique solution. This dissertation focuses on exemplar-based colorization methods, in which the extra information comes from a user-selected color reference image with similar semantic content to the target. While the user selects the reference based on content similarity, the algorithms estimate similarity based on local descriptors of image regions. This difference in abstraction between the user and algorithm perspective can lead to the algorithms not always being able to transfer colors between semantic corresponding elements in the image pair, specially in images in which the mapping between content/color and local descriptors is complex. Most exemplar-based methods in the literature display successful examples mostly limited to simple instances, such as landscapes, animals and simple buildings. Based on this observation, in this research we propose a new exemplar-based method that aims at generating plausible colorizations for a wider range of image pairs, including images of higher complexity. To that end, the proposed method features a two-stage classification scheme that uses the available features in a more consistent manner and makes the initial color assignments more robust. It also includes an edge-aware relabeling method that enhances the spatial coherence and mitigates the impact of the multimodality, inherent to the colorization problem, over the method's colorized outputs. In this dissertation, we present a broad review of the colorization literature introducing a taxonomy that categorizes colorization techniques based on the source of prior information used to guide their color assignments. The proposed method pipeline is then described in details, and its key modules are validated through experiments. Moreover, a comparative analysis is performed which subjects the proposed and baseline methods to different source/target pairs to visually assess and compare their results. Experimental results indicate that the proposed method yields colorization results that are more coherent and of higher visual quality compared to two state-of-the-art exemplar-based colorization algorithms, both in simple and complex image sets. The results also indicate that exemplar-based methods can achieve results of comparable visual aspect to those of modern deep learning approaches while allowing more user control.

Key-words: Image Colorization. Exemplar-based. Image enhancement.

[1] https://www.reddit.com/r/Colorization

RESUMO

A colorização de imagens consiste em, dada uma imagem em tons de cinza, gerar uma versão colorida plausível desta imagem, o que pode ser realizado como um processo manual/artístico[2] ou (semi-)automático. A colorização é um problema subdeterminado, sendo necessária informação extra para a obtenção de uma solução única. Esta dissertação foca em métodos de colorização baseados em exemplo, nos quais a informação extra vem de uma imagem de referência colorida e de conteúdo similar, selecionada pelo usuário. Enquanto a escolha da referência é baseada em similaridade de conteúdo, o algoritmo estima similaridades baseado em descritores locais. Esta diferença de nível de abstração faz com que tais métodos nem sempre sejam capazes de transferir cores entre elementos de semântica correspondente no par de imagens, sobretudo em imagens onde o mapeamento entre semântica/cores e descritores locais é complexo. A maioria dos métodos baseados em exemplo mostram resultados bem sucedidos em sua maioria limitados a imagens de mapeamento simples como paisagens, animais e construções simples. Nesta pesquisa, um método de colorização baseado em exemplos é proposto, com objetivo de criar colorizações plausíveis para um conjunto mais abrangente de pares de imagens de entrada do que os métodos existentes, especialmente imagens de maior complexidade. Para alcançar este objetivo, o método proposto conta com um mecanismo de classificação em duas etapas, que faz uso do conjunto de características extraídas das imagens de uma maneira mais eficiente. O método ainda inclui um mecanismo de refinamento da classificação inicial baseado nas bordas da imagem original, proporcionando maior coerência espacial ao resultado e ao mesmo tempo reduzindo o impacto da multimodalidade. Nesta dissertação, apresentamos uma revisão abrangente da literatura em colorização, introduzindo uma taxonomia unificada que categoriza as técnicas baseada na fonte de informação *a priori* utilizada para guiar a atribuição de cores. O método proposto é descrito em detalhes e seus principais componentes são validados experimentalmente. Além disso, uma análise experimental é realizada submetendo a técnica proposta e algoritmos selecionados da literatura à diferentes pares de imagens para avaliar e comparar visualmente os seus resultados. Os experimentos indicaram que o método proposto é capaz de gerar colorizações que são mais coerentes e de maior qualidade visual, em imagens simples e complexas, quando comparado com dois algoritmos baseados em exemplo do estado da arte. Os resultados obtidos também indicam que métodos baseados em exemplo são capazes de obter, em certas instâncias, resultados comparáveis aos dos novos algoritmos de aprendizagem profunda, enquanto permitem maior controle do usuário sobre o resultado.

Palavras-chaves: Colorização de Imagens. Métodos baseados em exemplo. Melhoramento de imagens.

[2] https://www.reddit.com/r/Colorization

LIST OF FIGURES

CONTENTS

1 INTRODUCTION

It is noticeable that colored pictures, photographs and images in general are more appealing to the human observer than grayscale ones, with humans being capable of discerning thousands of color shades and intensities while only dozens of shades of gray (GONZALEZ; WOODS, 2012).

Apart from the visual appeal aspect, the human response to colors also occurs in a cognitive level. The color content of an image is partly responsible for conveying information, being color one of the characteristics we use to recognize, describe and discriminate objects of interest (GORDON, 2004). Psychological studies indicate that color plays an important role in determining the gist of a scene from a quick observation, with the test subjects being able to classify a scene both quicker and more accurately when the scene is shown in its original colors compared to a gray version of itself or a fake color version (GOFFAUX et al., 2005).

The first application of *image colorization* came from the film industry, where black-and-white movies were colorized in a manual artistic process since the early 1900's. Computer-assisted colorization was introduced by Wilson Markle in the 70s also for the colorization of movies and was even used in images of the Apollo space program (COLORING..., 1986). The colorization process at that point was as simple as assigning predefined colors to gray intensities.

In the book by Gonzalez and Woods (GONZALEZ; WOODS, 2012) there is a section that defines pseudocolor image processing as the process of assigning colors to gray values based on a predefined criterion. Adding pseudocolors is useful for enhancing visualization and interpretation of images, not only in the aforementioned case of movies, where the objective is to mimic the actual colors of the original scene, but also in applications using different imaging equipment, where the colors are artificially generated for enhancing visual quality and interpretation, such as in scientific and medical imagery (MARTINEZ-ESCOBAR; FOO; WINER, 2012), thermal sensors (GU; HE; GU, 2017), night-vision (ZHENG; ESSOCK, 2008), radar (SONG; XU; JIN, 2018), infra-red (HAMAM; DORDEK; COHEN, 2012) and others.

Although the process of computer-assisted image colorization already existed in the form of pseudocolor image processing, this research focuses on the colorization algorithms that do not follow a predefined mapping from gray intensities to color values.

In 2002, inspired by the image analogies framework (HERTZMANN et al., 2001) and the color transfer technique (REINHARD et al., 2001), Welsh *et al.* (WELSH; ASHIKHMIN; MUELLER, 2002) proposed the first automatic colorization technique that does not rely on a predefined color mapping scheme. The algorithm receives a color source image and a grayscale target image and, based on intensity and statistical similarities between source and target, performs the color transfer from the former to the latter without the need of user assistance.

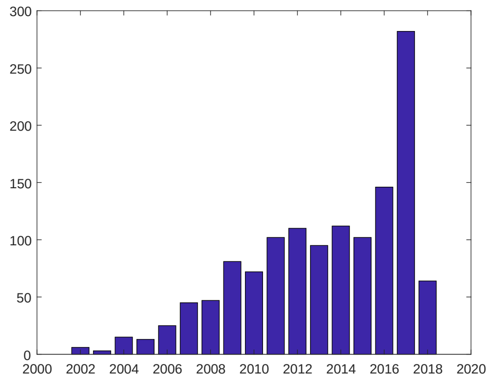


Figure 1 – Number of Google Scholar results for "image colorization" filtered by year. Until April 2018.

Since the pioneer work by Welsh, an increasing number of colorization methods have been proposed (Figure 1). There are some reasons that might explain this research interest. The most direct one is for the practical applications such as, colorizing old movies and photographs, automatic colorization of cartoons and mangas and also image compression (LEE et al., 2013). Furthermore, colorization resembles other problems in image processing/computer vision in which the goal is to predict values for each pixel of the input image using the image itself as a source of information while maintaining spatial coherence, such as predicting albedo, shading, depth and image denoising (DESHPANDE; ROCK; FORSYTH, 2015). Due to this shared structure, ideas and solution techniques might also be shared between them. There are still frameworks (such as (FATTAL, 2009) and (HUA et al., 2014)) that are directly applicable to colorization and other problems like detail enhancement, smoothing, etc. This shows that there is an intersection between colorization and other image processing problems that might be explored to create general solutions.

1.1 THE IMAGE COLORIZATION PROBLEM

The digital representation of a color image in most color systems consists of a matrix of 3 channels, with the color of a pixel being determined by the combination of the values of these channels. Meanwhile, a grayscale image is a single channel matrix in which each pixel consists of its intensity level. While the transformation of a color image to its grayscale counterpart can be as simple as computing a dot product of each color pixel with a set of values defined by convention, a single gray value can be mapped to many tuples, therefore the mapping between gray pixel to color pixel is not injective.

It would be straightforward to define the image colorization problem as the problem of finding the color image that generated its grayscale counterpart. In image compression, approaches that utilize colorization-based coding try to accomplish this, but they do it by saving part of the color information from the original image during the compression and then recolorizing the image based on this saved information during the decompression ((LEE et al., 2013), (BAIG; TORRESANI, 2017)). Without extra information, the colorization becomes underconstrained and therefore there is not a unique solution to the problem. To cope with this solution ambiguity, the colorization algorithms require prior information in order to generate a unique colorization result. The relationship between the prior information and the colorization methods will be addressed in the next chapter.

Provided the prior information, in image colorization the main objective is **not necessarily to recover the ground truth colors** of a grayscale image (which might not even exist) but to **add colors to this gray image in a coherent manner.**

There are two main aspects of this coherence that need to be addressed in order for the colorizations to seem visually plausible, in other words, for the colorized images to not look artificial. First, the colorization result must be semantically coherent. That means that the colors added must look realistic according to our notions of a real image, for example, if the clear sky is portrayed in an image, we expect the colors to be bluish (or even reddish if there is a sunset), but not green. The same idea applies to known objects or places. Second, the result must present spatial coherence, which means that pixel colors should be consistent with colors on the surroundings of this pixel. In the case of video colorization, there is also temporal coherence which means that the colorized images should be consistent with their neighbor frames, but we are not covering video in this dissertation.

1.2 PROBLEM STATEMENT AND RESEARCH OBJECTIVES

In this research we focus on exemplar-based colorization methods. In such methods, colors are transferred to the specified grayscale target image from a user-selected color reference image (see Section 2.1.2 for more details). The reference image is required to present similar semantic content to the target grayscale, so that the algorithm can explore these similarities to guide the color transfer process.

It was observed during the review of the exemplar-based methods ((WELSH; ASHIKHMIN; MUELLER, 2002), (IRONY; COHEN-OR; LISCHINSKI, 2005), (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008), (BUGEAU; TA, 2012), (GUPTA et al., 2012), (PIERRE et al., 2014), (LI; LAI; ROSIN, 2017)) that the reported successful results of such methods seem to be tied to simple imagery such as natural landscapes and images of single animals or buildings. These images are considered simple instances because the mapping between region colors and local descriptors is more straightforward in these images and therefore easier for the algorithm to grasp.

The observation of the scope limitation of the available exemplar-based methods leads to the question of whether it is possible to successfully apply such methods to different and more complex types of input images. To assess this question, we propose a new exemplar-based colorization method, designed to handle a broader set of input image pairs than the existing exemplar-based techniques.

In order to evaluate the effectiveness of the proposed technique in dealing with this wider scope of images, we subject the method along with two state-of-the-art exemplar-based methods to a selected set of input image pairs and compare the results generated by each. The selected pairs are composed of both images considered simple and complex. The result assessment and comparative analysis are performed through visual inspection, observing how well the algorithms are able to transfer colors between corresponding elements of reference and target, if the transferred colors observe the original structure of the gray image, if the output presents spatial coherence, among other aspects. By visually examining the colorization generated by the proposed method we can assess whether an exemplar-based technique is able to successfully work around the aforementioned limitations.

1.3 DISSERTATION STRUCTURE

The remaining of this document is structured as follows. In Chapter 2, a broad literature review on digital image colorization is presented. The review features a proposed unified taxonomy that categorizes each method according to their source of semantic *prior* information and presents important algorithms from each category. The chapter ends summarizing the presented review, discussing limitations of current methods and research trends.

With the limitations of current methods considered in Chapter 2, in Chapter 3 we describe the proposed colorization method. Each component in the method's pipeline is detailed, and the motivation behind the design decisions are explained.

Then, in Chapter 4, the experimental portion of this research is covered. The evaluation method is described and experimental results are presented to both validate the proposed method's design and compare the final results to state-of-the-art methods.

Finally, Chapter 5 outlines the main points throughout this dissertation, including a summarization of the conclusions derived from the experiments. The chapter ends with possible directions for future works.

2 LITERATURE REVIEW

Since the pioneering work by Welsh (WELSH; ASHIKHMIN; MUELLER, 2002), many algorithms have been proposed with the intent of tackling the colorization problem. Colorization methods in the literature can vary greatly both in terms of how they approach the problem, with different types of input information and theoretical frameworks, and in terms of their solution techniques.

In this chapter, we present a review of the published research in image colorization, tracing from the early stages to the most recent works, while also making an effort to categorize the presented methods and point out research trends.

2.1 CLASSIFICATION OF COLORIZATION TECHNIQUES

As mentioned in Chapter 1, colorization techniques require a source of prior information to resolve the ambiguity inherent to the problem. According to the source of prior information, colorization techniques were originally classified in most of the literature into two major groups: *scribble-based* and *exemplar-based* methods.

In scribble-based methods, the prior is composed of color annotations (scribbles) drawn by the user directly onto the target grayscale image. The algorithm is then responsible for propagating this color scribbles to the rest of the image in a coherent manner in order to complete the colorization. On the other hand, in exemplar-based methods, the prior comes in the form of a color image (or set of images) with similar semantic content to the target image. The algorithm is responsible for transferring the color from the reference (*source*) image to the grayscale (*target*) image.

Since the introduction of methods that utilize datasets of images for colorization, the aforementioned two classes does not cover the whole spectrum of techniques. The taxonomies presented in the current literature vary among papers, so we introduce our own working classification in this review in order to avoid any confusion. The classification is presented in Figure 2.

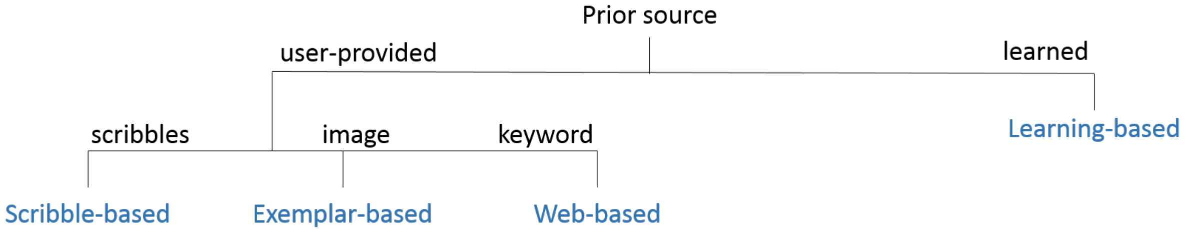


Figure 2 – Classification of colorization techniques according to the source of semantic prior information.

The main classification criterion we use is the type of semantic prior information used by the algorithm. In this context, semantic information means information related to the interpretation of the scene (*scene understanding*). The semantic prior can be provided by the user in different ways (left branch in Figure 2) or automatically learned from data (right branch). We classify the algorithms, as shown in the figure, according to the way the semantic information is conveyed to them. This also happens to be a measure of the amount of user interaction required from each algorithm category.

In the remainder of this chapter, we will cover each algorithm category by presenting the most important research exemplars.

2.1.1 Scribble-based methods

In scribble-based methods, as previously described, the user “paints" the desired colors directly onto portions of the target image and the algorithm then propagate the colors to the rest of the image to complete the colorization. The process is illustrated in Figure 3. On the left side the input gray image is shown with overlaid user scribbles, on the right the output of the scribble-based colorization algorithm is shown.

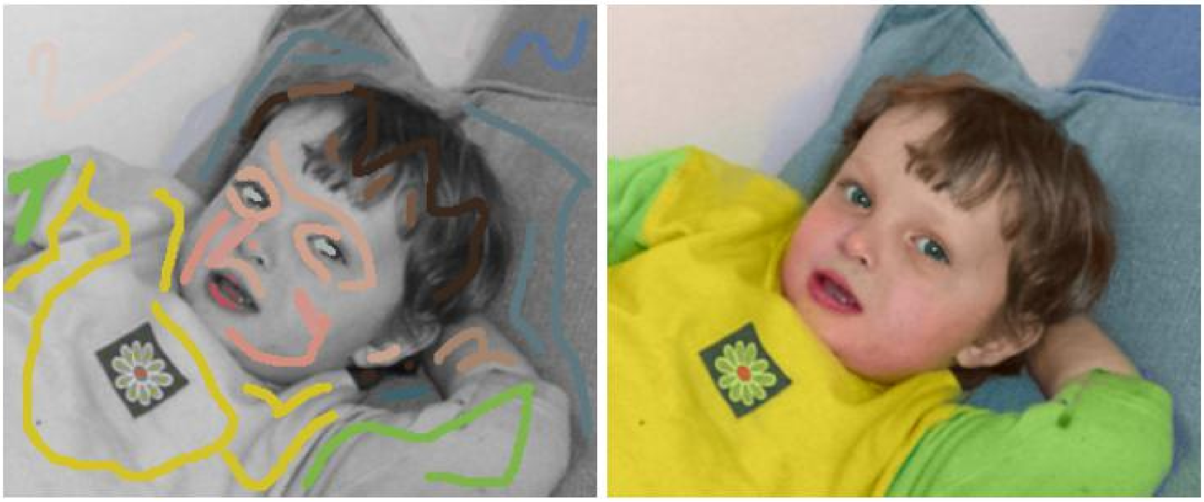


Figure 3 – Example of Scribble-based colorization (Source: (LEVIN; LISCHINSKI; WEISS, 2004))

Since the user provides the colors directly onto the image, the semantic aspect of the colorization (discussed in Section 1.1) is handled by the user during the color casting and placement. Also, as the input is provided manually, it requires a certain degree of knowledge from the user about the algorithm behavior and in many cases becomes an iterative process of drawing inputs, evaluating the colorization result and repeating until it meets the desired goal.

In their seminal work, Levin *et al.* (LEVIN; LISCHINSKI; WEISS, 2004) proposed the first scribble-based colorization method. The algorithm is based on the premise that nearby pixels on the image that have similar gray levels should also have similar colors. They translate this premise into a global optimization problem that minimizes the difference between a pixel chrominance and the weighted average of its neighbors. The optimization variables are the pixels intensities in each chrominance channel and the weights are a function of the difference of the pixel intensity values. The user scribbles are used as constraints so the algorithm changes the chrominance values around the image while also preserving the initial color assignments. Due to the nature of the cost function that minimizes differences around neighborhoods, this method introduced the idea of spatial coherence, which was not present in the previous works.

While Levin's algorithm enforces similar colors for similar intensities, boundary regions, which present abrupt intensity changes, tend to become blurred due to different colors propagating and interlacing to the same area, generating color bleeding artifacts. This usually requires extra effort from the user to carefully place scribbles around complex boundaries. Huang *et al.* (HUANG et al., 2005) proposed an enhancement of the previous technique by generating an edge map from the gray image and integrating this edge-map into the color propagation step in order to avoid blurring around the edge regions.

To overcome the computational cost of the global optimization used in the color propagation techniques, Yatziv *et al.* (YATZIV; SAPIRO, 2006) proposed an scribble-based method that is solved iteratively. Each pixel is considered as a structure that contains its gray intensity value along with a list of chrominances. The list of chrominances of each pixel is filled iteratively by sharing colors of *linked* pixels in a process that resembles a graph search. With the lists of chrominances of each pixel, the final colors are determined by applying a color blending technique. The visual comparison shown in the paper indicates that they achieve result of quality similar to their predecessor but in only a small fraction of the time.

Another drawback of the aforementioned works is that they rely on low-level similarity (pixels distances and intensity differences) to perform the color propagation. Therefore, in image regions that are rich in details, these techniques require many scribbles to provide the desired results. To overcome this issue Qu *et al.* (QU; WONG; HENG, 2006), Luan *et al.* (LUAN et al., 2007), and Sheng *et al.* (SHENG et al., 2011) designed similarity metrics that utilize image pattern continuity to increment the low-level similarity during propagation.

Qu dealt with manga colorization, which possess many strokes and drawing techniques (such as hatching and screening). Such techniques generate many intensity discontinuities which would require an unmanageable amount of scribbles from the previous techniques. The algorithm utilizes a Gabor filter bank to generate pattern/texture features and propagate colors inside regions with similar patterns according to a clustering of these features. Sheng also used Gabor filter banks for texture description in colorization. The algorithm generates a rotational invariant feature by adapting the Gabor filter banks and then for each pixel, it computes the *interimage neighborhood* which consists of the pixels neighbors in the texture feature space. The algorithm then performs color transfer/propagation through an optimization similar to the one in (LEVIN; LISCHINSKI; WEISS, 2004) using the more general interimage neighborhood instead of the pixel neighborhood. The technique was applied both to drawings and natural images and also supports exemplar-based colorization.

Luan, on the other hand, targeted natural images, which are characterized by "rich and inhomogeneous texture distributions". The technique is divided in two steps: *color labeling* and *color mapping*. The user first draw scribbles in order to group similar regions in the image. Since the technique analyzes texture similarities in a non-local fashion, the user does not need to cover the entire image with scribbles, having only to cover a small subset of each region of interest. The color labeling step group similar regions throughout the whole image, it does so through the optimization of an energy-based cost function that is composed of an intensity continuity term and a texture similarity term. For each pixel, the weights of each cost term are determined by the smoothness of the region around the pixel, which favors the intensity term for pixels within smooth regions and the texture term in non-smooth regions. Then, in the mapping step, the user selects a few pixels from each region to assign colors and the remaining pixels in each region are colorized by linear interpolation based on their intensity levels and the color of the reference pixels. A post-processing step applies color blending to make the color transition more natural in boundary regions.

Balinski *et al.* (BALINSKY; MOHAMMAD, 2009) used a Bayesian analysis to tackle the colorization problem. The authors decided to study the distribution of the response to the low-pass neighborhood filter proposed in (LEVIN; LISCHINSKI; WEISS, 2004) applied to the chrominance channels of natural images and found empirically that they belong to a heavy-tailed non-gaussian distribution. They proceed to use the filter response as a regularization term in the Bayesian setting which yields a generally non-convex optimization problem that is the generalization of the cost proposed in (LEVIN; LISCHINSKI; WEISS, 2004). To convexify the problem the authors opt for the L^1 norm and solve it through a linear program. The results are visually similar to Levin's when the inputs are scribbles, but when the inputs become sparse (color pixel seeds) they produce more vivid colors with more well defined boundaries than the L^2 counterpart. One drawback of the approach is the optimization technique for the L^1 norm which is significantly slower compared to least squares closed-form solution.

2.1.2 Exemplar-based methods

In exemplar-based methods, the user provides a color source image with similar content to the gray target image as prior information. The algorithms then explore pattern similarities between the images to transfer colors from source to target, as shown in Figure 4.

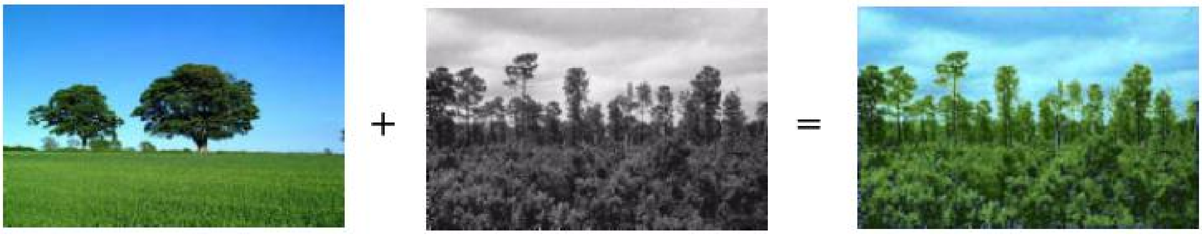


Figure 4 – Example of Exemplar-based colorization (Source: (WELSH; ASHIKHMIN; MUELLER, 2002))

The semantic coherence is again of user responsibility, this time through the choice of the reference image. Since the user does not work directly onto the target image, the approach is more automated than the scribble-based counterpart, but, the result quality becomes limited by the availability of the source images.

To assign colors for each target pixel, the majority of the exemplar-based algorithms make use of pixel neighborhoods, because single pixel values do not carry enough information. Instead of using pixel neighborhoods to propagate input colors (as the algorithms in 2.1.1), these methods use them to build local descriptions of the pixel for further matching between source and target. These exemplar-based methods are part of the *patch-based* image processing techniques (BUGEAU, 2018). Patch-based techniques rely on the principle of self-similarity which states that natural images present some level of predictability or redundancy within themselves (HYVÄRINEN; HURRI; HOYER, 2009). Exemplar-based methods extrapolate this concept by expecting the redundancy to be present not only within an image but also between images of similar content as well.

In the pioneer work by Welsh *et al.* (WELSH; ASHIKHMIN; MUELLER, 2002) a simple exemplar-method that leverages on simple neighborhood statistics is proposed. First, the images undergo pre-processing which consists of color space transform (RGB to Lab) and luminance remapping, so that the luminance channels of both images become comparable. Following the pre-processing, source sampling is performed to reduce the computational cost of the following steps. Then, for each pixel in the target and for all samples in the source image a feature vector of two dimensions is computed composed of the luminance of the pixel and the standard deviation of the window centered at the pixel. For each target pixel, the algorithm assigns the color of the source sample that is the closest in feature space to the target. The presented results are satisfactory on very simple images, and the authors claim that the method works well on scenes where the images are divided into luminance clusters or have distinct textures.The authors also propose a workaround for more difficult scenes by asking the user to place rectangles (called swatches) in both source and target to confine the color transfer to the pixels that belong to the swatches and then transferring for the remaining of the target based on the already colorized swatches.

One of the main drawback of Welsh's work is that it uses a greedy approach to colorize each pixel with its nearest neighbor in feature space and therefore does not encourage spatial coherence in the target image. Irony *et al.* (IRONY; COHEN-OR; LISCHINSKI, 2005) took advantage of scribble-based techniques' strengths to enforce spatial coherence in an exemplar-based framework. The premise of their work is that good pixel/neighborhood matches between two images are not enough for a successful colorization, because matching images might come from different contexts and therefore possess different colors. Instead of relying only on low level information such as intensity and neighborhood statistics, the algorithm requires a partially segmented source image as input. The algorithm extracts local features using a windowed dct and performs custom-tailored dimensionality reduction to generate a low-dimensional space where it performs a classification of each target pixel to one of the source segmented regions. To enforce spatial coherence, the method includes an image space voting technique that assigns confidence levels to the labeling of pixels based on its accordance with neighborhood labels. Finally it transfer colors to pixels based on a weighted average of predictions from the pixel neighborhood, but instead of assigning the pixel colors directly, it assigns only for those with high confidence and uses them as input *micro-scribbles* to colorize with the algorithm from (LEVIN; LISCHINSKI; WEISS, 2004).

In (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008) the authors state that choosing colors based solely on the local description of pixels is prone to error because different color regions might present similar local description (*multimodality*) in a source image. Instead of deciding pixel colors at the local level, their model first estimates the probabilities of each color for each pixel and then performs the color decisions in a global level based on these probabilities. The algorithm first discretizes the source image color space, creating classes to represent color ranges, favoring regions with more density of observations. It also generates a local descriptor for each pixel composed of dense surf, intensity value, and windowed standard deviation and Laplacian. Then, based on the observations (pixels) from the color source, it computes for each pixel an estimate of the conditional probability distribution of *color given descriptor* using Parzen windows. The prior distribution of descriptors is also computed based on the observations. For the final color assignment, it applies graph-cuts to optimize an energy cost that maximizes the posterior of the aforementioned probabilities with a dedicated term for spatial coherence.

Gupta *et al.* (GUPTA et al., 2012) explored superpixel resolution to speed up a complex colorization pipeline and encourage spatial coherence. In the first step, both source and target images are segmented into superpixels through a geometric-flow based algorithm. Then each superpixel is assigned a feature vector containing average intensity, standard deviation, Gabor and surf features. To reduce computational cost, the feature matching is computed in a cascade fashion that iteratively prunes the search space using each feature separately and then the final matching utilizes the whole set of features to transfer color to the pixel at the center of each superpixel (micro-scribbles) from their corresponding nearest neighbors. The color of the remaining pixels are obtained using (LEVIN; LISCHINSKI;

WEISS, 2004) to propagate the assigned micro-scribbles (such as in (IRONY; COHEN-OR; LISCHINSKI, 2005)). The final step consists of a color reassignment that segments the image once again, in a coarser scale, and analyzes the clustering of superpixel colors within each new segment. It considers that dense clusters represent high confidence color assignments and performs color reassignment otherwise. Visual comparisons indicate that this algorithm outperforms (WELSH; ASHIKHMIN; MUELLER, 2002), (IRONY; COHEN-OR; LISCHINSKI, 2005) and (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008) with a considerable improvement over them in the presented scenes.

In a series of works, Bugeau and Pierre *et al.* applied the primal-dual algorithm (CHAMBOLLE; POCK, 2011) to solve the colorization problem in a variational framework (POPURI, 2010). In (BUGEAU; TA, 2012), the authors start by extracting local features from each pixel in the image (variance, amplitude spectrum of dft and luminance histogram were used). For each feature a simple distance is designed, and the combination of these distances generates a metric in feature space. This metric is used to select, for each pixel, a group of source candidates (closest in feature space) and then the median of the candidates chrominance is used for color transfer. As the chrominance candidates set is two dimensional, the median of the set is defined as the element that is projected to the median of the first component of the pca of the set. To enforce spatial coherence a post processing step is proposed with an optimization based on tv regularization. Then, in (BUGEAU; TA; PAPADAKIS, 2014) the proposed method utilizes the same feature extraction and candidate selection steps, but the color selection and tv regularization are performed in a single optimization which makes the process more reliable since the coherence constraints are enforced during the color assignment itself. The results in both papers show good level of spatial coherence, but the generated color images seem washed-out presenting desaturated/bland colors. These bland colors seem to be a byproduct of the regularization which leads to a tradeoff between spatial coherence and color diversity. In (PIERRE et al., 2014) the authors suggest that better color consistency can be achieved by working directly in the RGB color space as opposed to the vast majority of the related research. The authors formulate an optimization problem over the RGB space and therefore colors are assigned directly onto the three RGB channels. The algorithm presented an improvement over (BUGEAU; TA; PAPADAKIS, 2014) and provide better visual results than (WELSH; ASHIKHMIN; MUELLER, 2002), (IRONY; COHEN-OR; LISCHINSKI, 2005) and (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008) in the presented images. But in comparison with (GUPTA et al., 2012), Gupta's work still presented more vivid and overall better colorizations. Finally in (PIERRE et al., 2015) the authors design a model that couples the channels in the YUV color space and provide an optimization algorithm with proof of convergence. The model supports the use of different priors (examples and scribbles) and works particularly well with images that present thin structures outperforming (GUPTA et al., 2012) in this regard.

More recently, Li *et al.* (LI; LAI; ROSIN, 2017) focused on local feature selection to enhance colorization results. The paper proposes an automatic feature selection method that uses intensity features for uniform regions and texture features for regions considered non-uniform. One of the main contributions is the automatic classification of image regions (superpixels) that first estimates the probability of each region belonging to one of the two classes using a Bayesian inference scheme and then solving the classification problem through the optimization of a mrf cost function. The cost considers not only the isolated probability of a superpixel belonging to a certain class, but also the interactions with the superpixel's neighborhood labels to enforce spatial coherence. To further guarantee consistency, the image is segmented into superpixels in a coarser scale and region labels are reassigned according to majority voting. The same author went in a different direction in (LI et al., 2017), posing the colorization as a dictionary-based sparse representation problem. The images are first segmented into superpixels and each superpixel has a feature vector assigned to it that combines low (intensity related), mid (DAISY descriptor) and high level (saliency detection) features. The set of features of the reference image is used as the dictionary and the chrominance transfer becomes an instance of sparse matching. The authors also include a locality consistent regularization term in the cost that favor target superpixels that are close both in feature and image space to have similar dictionary representations, hence promoting spatial coherence during the matching itself.

2.1.3 Web-based methods

Web-based methods are similar to exemplar-based methods in the sense that they also use example images to perform the colorization. But, instead of relying on images provided directly by the user, the web-based methods leverage on the large amount of images available on the internet to reduce user interaction. The user provides an input to an internet search (usually a keyword) and the system chooses among the set of results the ones that are adequate to be source images for the target at hand. The semantic information provided by the user is the search input and there is no need for image selection, therefore, this class of algorithms are more independent than the previous, but at the same time, the step of source selection/filtering can be quite expensive.

In (LIU et al., 2008) the authors propose a system that works very well with images of famous monuments and places or rigid structures/buildings. They focus on how to solve the illumination inconsistency between source and target that might occur when multiple shots of a same scene are taken under different conditions. To achieve a colorization that is invariant to illumination changes, the authors use multiple internet images to generate what they call the intrinsic reflectance image. The set of images from the search results are registered to the target by matching sift features and, depending on the matching error, the registration is carried via global alignment or triangle-based warping. From the set of registered images it is possible to perform illumination/reflectance decompositions (for color and gray versions) and therefore generate illumination-independent images. The color transfer is performed directly from color reflectance to target reflectance on pixels with small registration errors and used as color seeds (micro-scribbles) to the algorithm in (LEVIN; LISCHINSKI; WEISS, 2004). Due to the robustness provided by the multiple images, the algorithm outperforms other exemplar-based methods in scenes of famous locations, but, due to the requirement of registered versions of the same image, it should only work in these scenarios.

Morimoto *et al.* (MORIMOTO; TAGUCHI; NAEMURA, 2009) utilizes a very large scale search to provide an entirely automatic colorization approach. Their approach is different in the sense that it does not utilize any source from the user, instead it gathers a very large set of images (1 million) from the web and filters to 100 results based on global image descriptors (ssd of the *gist* scene descriptors). Based on the filtered results, the algorithm produces multiple colorized images utilizing an algorithm very similar to Welsh's to produce each individual result. The algorithm seems very dependent upon the gist matching and considerably expensive to compute in such a high volume of images, while only presenting a single result.

Chia *et al.* (CHIA et al., 2011) proposed a system that generates multiple candidate colorizations for a given target. First, the algorithm requires the user input which consists in the target image with segmentation of important foreground objects and semantic labels for each segmented object. The first step of the algorithm consists of downloading a set of images from the internet for each foreground object (approximately 30 thousand) using the provided labels and then utilizing *saliency detection* to obtain the foreground elements from each search results. It then applies contour consistency to filter objects from the search according to shape similarity compared to the target objects. For object selection, the foreground objects undergo feature extraction utilizing intensity, Gabor wavelets and sift features and the background are compared in terms of their gist descriptors. The best internet references are used to provide diverse colorization possibilities for the foreground objects. Color assignment is carried in superpixel level through global optimization of an energy function that accounts both the distance in feature space and the smoothness of the colorization result using the belief propagation framework. The results are compared to (WELSH; ASHIKHMIN; MUELLER, 2002), (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008) and (TAI; JIA; TANG, 2005) and show more convincing colorizations confirmed by a user study with the results being labeled as real in up to around 66% of cases. The downsides are the need of object segmentation and, although not discussed in the paper, there is probably a high computational cost involved.

Wang *et al.* (WANG et al., 2012) proposed a system for *affective* image colorization in which the user provides an affective word that is used for the selection of color themes based on art theories. Initially, the target image is semi-automatically segmented by a graph-cut technique and, such as in (CHIA et al., 2011), the user provides labels for foreground objects and the results are filtered according to saliency detection and contour consistency. Color themes are mapped to affective words through a coordinate system using the lasso regression framework (O'DONOVAN; AGARWALA; HERTZMANN, 2011). Then an energy-based optimization cost is designed to select the best reference object for each target object according to both the consistency with the target and consistency with the given emotion defined by the word. After source selection, the color transfer is performed as another energy-based optimization.

2.1.4 Learning-based methods

The last class of methods are the learning-based in which the user provides the colorization model with a dataset of color images for training. Once the model is trained, it is capable of colorizing a previously unseen image, in most cases without requiring any additional information. Although the web-based methods also utilize large sets of data, in learning-based methods the whole set of inputs is used to perform an *offline, one time* training step, in contrast to web-based which perform a new web search plus filtering for every new input image.

But, the main difference that sets this class of methods apart from the aforementioned classes is that the semantic information does not come from the user, which allows the methods to be completely automatic.

In (CHENG; YANG; SHENG, 2015), Cheng *et al.* proposed the first deep learning algorithm for image colorization. The authors argue that high-level understanding of an image can be useful to perform low-level vision tasks therefore they propose a feature descriptor that includes semantics into the colorization. The approach utilizes a deep neural network that takes as input feature vectors and output the chrominance values (in YUV color space) for each pixel. The feature vector structure is composed of three parts: low-level (intensity window around pixel), medium-level (DAISY descriptor), high-level (semantic label obtained by a scene parsing algorithm). During the training stage, training pairs composed of the feature vector and the chrominance of a pixel are fed to the network that utilizes a least squares regression framework that compares predicted chrominance output with the ground-truth. After the training stage is completed, the feature vectors of each target pixel can be fed to the network which in turn outputs the initial chrominance values. The initial chrominance values are refined by applying a joint bilateral filtering to remove artifacts and guarantee spatial coherence. The results shown in the paper, even for training on a relatively small set (around 2.7k images), indicate that the algorithm is able to outperform the state of the art algorithm from Gupta *et al.* (GUPTA et al., 2012) both visually and in terms of target image psnr, at least in the chosen set of examples.

Although (CHENG; YANG; SHENG, 2015) features a learning-based approach, it still works with a small training set that does not contemplate a diverse set of scenes. In the concurrent works presented in (IIZUKA; SIMO-SERRA; ISHIKAWA, 2016), (LARSSON; MAIRE;

SHAKHNAROVICH, 2016) and (ZHANG; ISOLA; EFROS, 2016), the authors make use of cnn and manage to scale the training to the order of millions of images.

In (IIZUKA; SIMO-SERRA; ISHIKAWA, 2016), Iizuka *et al.* utilized a cnn that combines both local information obtained from image patches with global priors obtained from the image as a whole. Although they also consider the features as blocks of low, medium and high level, unlike the work in (CHENG; YANG; SHENG, 2015), there are no hand-crafted feature descriptors. The network weights themselves are interpreted as a volume of features, as if each layer of the network sees a filtered version of the image. The proposed architecture contains two branches, one for obtaining the global level features and other to fuse this global information with the remaining features and then generating the output. The high level (global) features act as semantic priors that indicate the type of image (indoor vs. outdoor, day vs. night, etc.) so that the local features chooses colors more appropriate to the scene being colorized. The training is performed using the back-propagation algorithm with a simple mse cost function. To improve the final results, the authors train a small scene classification network jointly with the full colorization network, the errors of the classification net are *backpropagated* to the global features network to help guiding its optimization. After training on a dataset of about 2.5 million images, the algorithm is able to colorize (in less than a second) a very diverse set of images, including even indoor images, images containing people and legacy grayscale images which were not much explored in previous works. In the user study, the output of the model is considered “natural” by the user in more than 90% of the images.

Larsson *et al.* (LARSSON; MAIRE; SHAKHNAROVICH, 2016) proposed another fully automatic system with two design considerations in mind: the need to include semantic information in the pipeline and the need to model the multimodality. To tackle the multimodality, instead of designing a loss function based on color differences, the authors estimate color distributions for each pixel (such as the idea in (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008)) and utilize a kl loss function to compare the estimated distribution against the ground-truth, which consists of the distribution of the window around the pixel. To generate the final color from the histogram predictions, the authors propose different inference mechanisms and state that for the Lab color space the best qualitative and quantitative results are achieved through computing expectations weighted by the histogram values. The authors propose a benchmark for future research using learning-based methods. It is composed of 10k images from the ImageNet dataset (DENG et al., 2009) with a balanced representation for the set categories and evaluation through rmse and psnr.

Zhang *et al.* (ZHANG; ISOLA; EFROS, 2016) proposed a system that focuses on achieving colorizations that are visually plausible but at the same time present diverse and vivid colors. The authors argue that the colorizations from previous learning-based works look desaturated because of their use of standard L^2 regression loss functions that lead to conservative predictions (as pointed out in (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008)), so, in order to produce more vivid colorizations, the authors propose a classification loss that compares color probabilities using cross entropy. The authors presented empirical evidence that the distribution of colors in natural images is highly biased towards desaturated colors and so they proposed a class rebalancing scheme to encourage colorful colorizations. After training on over a million images, the authors performed a large scale user study using the Amazon Mechanical Turk and show that the method is able to fool humans in a Turing-like test on 32% of trials.

Deshpande *et al.* (DESHPANDE et al., 2016) also targeted the multimodality of colorization but with the intent of creating multiple coherent colorizations for the same gray image. The authors propose to learn the conditional distribution of chrominances given a gray image so sampling this distribution would allow to obtain different colorizations. Given the large amount of combinations of colors and images the distribution becomes too scattered and therefore it is necessary to find a low dimensional representation of the chrominance space. The authors propose the use of vae to obtain a smooth low-dimensional (encoded) representation of the chrominance span and an efficient decoder that generates a plausible colorization for any given point in this low-dimensional space. The decoder loss function is designed to enforce spatial and semantic coherence as well as colorfulness as in (ZHANG; ISOLA; EFROS, 2016). To model the conditional probability of an element of the low-dimensional subspace given the gray image, the system uses a mdn since it allows for the output vector to take many values given the same image, providing the desired diversity. The results show that the algorithm is able to produce multiple plausible results which might be useful to present options for posterior user evaluation.

Zhang and Zhu *et al.* (ZHANG et al., 2017) proposed a deep learning approach for user guided colorization where the system takes as input from the user sparse color annotations over the target image and generates a complete colorization. Unlike the propagation algorithms discussed in Section 2.1.1, this system end-to-end learns the mapping between gray image plus color seeds to fully colorized images without relying on hand-crafted rules for the propagation. Since the user is able to directly resolve the color ambiguities by placing the color seeds, the system does not need to account for multimodality and therefore a simple regression loss is applied, the authors choose the L^1 norm. To be able to generate a large scale training (over a million images) without requiring user provided color seeds for these many images, the authors simulate user inputs by sampling the original color image and providing the sampled colors as seeds. To evaluate the propagation of samples, the authors utilize the psnr and show that for a few seeds the algorithm outperforms (LEVIN; LISCHINSKI; WEISS, 2004), but once the number of seeds grows to a few hundreds both algorithms perform the same which indicates that as the number of seeds grow the mid to high level information learned by the model becomes less important and the low-level optimization from (LEVIN; LISCHINSKI; WEISS, 2004) becomes enough to solve the problem. The system is equipped with an user interface that provides color suggestions in real time and it is reported that users are able to produce realistic colorizations in less than a minute.

2.2 DISCUSSION

In the early colorization methods, we notice a trade-off between the quality of results and the amount of user input required by the algorithms. Scribble-based techniques demonstrated more potential to generate the most realistic colorizations compared to the exemplar-based methods provided that the user has knowledge about the process and iterate until achieving the desired results. However, the more automatic approaches are interesting due to their potential to create results with less user effort and for a deeper understanding of the problem intricacies, which can be translated to similar problems.

It seems that the main research interest in the subject headed towards more automatic data driven approaches, a phenomenon we can observe also in many other computer vision research topics as well. Initially with the web-based approaches that leveraged on the easy access to large set of images on the web, to the current state-of-the-art learning-based methods that take advantage of the high computational power and large image datasets available to train their cnns. cnns in particular seem to be the current research focus, specially because for the colorization setting, they require little to no data pre-processing while allowing for fully automatic algorithms that do not require hand-crafted filters or descriptors.

The presented literature review indicated that semantic prior information is key for colorization. In scribble-based methods, since the user is responsible for providing colors directly onto the target, the algorithms can rely solely on low level image features (such as intensity and simple statistics) to propagate the user scribbles. On the other hand, for the remaining classes, the color information comes from example images and therefore the most robust algorithms seem to be the ones that are able to extract and integrate image features of different levels. The learning-based methods for example have no user-provided semantic information and therefore need to extract high level information during training in order to present convincing results. Some learning methods even include a dedicated network to extract high level information to pass to the colorization main network. The capacity of extracting high level information allows for successful colorization instances in complex scenery including artificial objects, buildings and people. The surveyed exemplar-based algorithms on the other hand, do not incorporate high level information into their pipelines. Since the user selects reference images based on content similarity and the algorithm measure similarities based on local descriptors of low to medium level, the algorithms might not be able to transfer colors correctly between elements of similar content if the intensity and texture information is not enough to discriminate these image elements. This might be one of the reasons for the results presented by these algorithms being mostly on simple images such as nature landscapes and animals.

Regarding the multimodality, since similar local descriptions of pixels might come from regions of different semantics and colors, the correct matching of the local descriptors in exemplar-based methods is not enough for an accurate color assignment. This issue might go unnoticed if the selected input pair does not possess distinct regions with similar descriptions, but otherwise, color assignments based solely on local decisions are prone to errors. Since learning-based methods need solutions that scale for large datasets with images of different characteristics, explicit treatment of the multimodality is required and was included in the most successful methods. However, even though the multimodality issue was initially presented in an exemplar-based framework (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008), the later representatives of this class did not directly targeted this issue in their pipeline, relying on color post-processing to rectify initial errors. This might be another reason for most of the favorable results of the surveyed exemplar-based algorithms to be on simple images.

As for result evaluation of image colorization methods, since the colorization objective consists of generating a plausible color version of a grayscale image and not necessarily recover original colors, comparisons with ground-truth most often not considered. For the majority of works in the literature, the evaluation is performed through visual inspection. The visual evaluation allows to assess the result images coherence and overall quality, however, it is a subjective criterion that depends on the observer's judgment. To enhance the results analysis, some works in the literature elaborate user studies which consists in visual inspection performed by a group of users with the intent of deriving statistics, making the result less subjective.

Based on the observation of the limitation of the available exemplar-based methods and its possible causes discussed in this section, in the next chapter we propose a new exemplar-based technique. This technique aims at generating plausible results for a broader set of input images by designing modules that deal with the limitation of local descriptor representations.

2.2.1 Exemplar-based methods and Scene Complexity

Exemplar-based methods are characterized by transferring color from an user-selected color reference image (*source*) to the *target* gray image. The algorithms compute local descriptors for each pixel on both source and target images and then based on distances between the source and target descriptors transfers colors from the former to the latter.

While the algorithms utilize **local descriptors** to establish relationships between source and target elements, the user selects the reference based on **semantic content similarity**. This difference in how the user and the algorithm perceive the image similarities can lead to unexpected results.

First, if the elements of similar semantics from source and target generate dissimilar local descriptions, it is hard for the algorithm to correctly associate the corresponding elements of these images. We call this the semantic *multirepresentability* issue. Second, if elements of similar local descriptors in the source image possess distinct colors, even if the algorithm is able to correctly match target to source descriptors, the color assignments will be misguided. This phenomenon was already observed (IRONY; COHEN-OR; LISCHINSKI, 2005) and defined as the *multimodality* effect (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008). Therefore, our notion of complexity of an input pair is related to the disposition of the image elements (pixels/superpixels) in the descriptor space.

Visual characteristics of the images in the input pairs may trigger the aforementioned issues or present other challenges to colorization algorithms. For instance, low resolution or blur in the images can cause the local descriptors to not be able to create distinctive representations and therefore can cause both of the above issues. Color imbalances in the source can introduce bias in the target color assignments while too many colors in the source can make the descriptor space too cluttered.

To be able to create plausible colorization for more complex input pairs, the method must include robustness elements that account for the possibility of local classification/-color assignment errors due to local descriptor limitations.

3 PROPOSED METHOD

As discussed in Section 2.2, most exemplar-based methods in the literature seem to have their successful colorizations limited to simple input images, such as natural landscapes and animals. These images are considered simple because they are usually divided into regions of distinct textures and these regions usually possess different color tones. The presence of these clearly distinguishable regions causes the mapping between colors and local descriptions of pixels to be easier for the algorithm to establish.

With that in mind, we propose an exemplar-based method that aims at generating plausible colorizations for a broader set of input images. The method works at superpixel level which enhance the coherence of mappings, and combines different features to create a descriptor space in which metrics are more distinctive. It also introduces a two-stage classification which improves the accuracy of target superpixels color labeling, and an edge-aware relabeling scheme that enforces spatial coherence while respecting the original image structure.

The remaining of this chapter is organized as follows. In Section 2.2.1, we explain the complexities faced by the exemplar-based methods in general. Then in Section 3.1 we present our method pipeline, describing each of its modules in details. Finally Section 3.2 presents some afterthoughts on the implementation of the method.

3.1 METHOD PIPELINE

Figure 5 shows a simplified abstraction of the proposed method's pipeline. In the following subsections, we detail each step in this pipeline.

3.1.1 Image Preprocessing

The method starts with a preprocessing stage that prepares the input images to the following stages of the algorithm. The preprocessing consists of a color space transformation and a histogram manipulation step.

In exemplar-based colorization, the methods rely on transferring only the color information from the source image to combine with an already existing luminance channel at the target image. Therefore, the source image should be transformed to a color space in which intensity and color (chrominance) information are decorrelated so the algorithm can perform the transfer operations in the chrominance channels without affecting the intensities. In (REINHARD; CUNNINGHAM; POULI, 2013), the Lab color space presented the smallest channels covariances amongst the surveyed spaces, and therefore it is a suitable choice for the proposed method. The proposed algorithm then starts by transforming the input source image from the conventional RGB representation to the Lab color space.

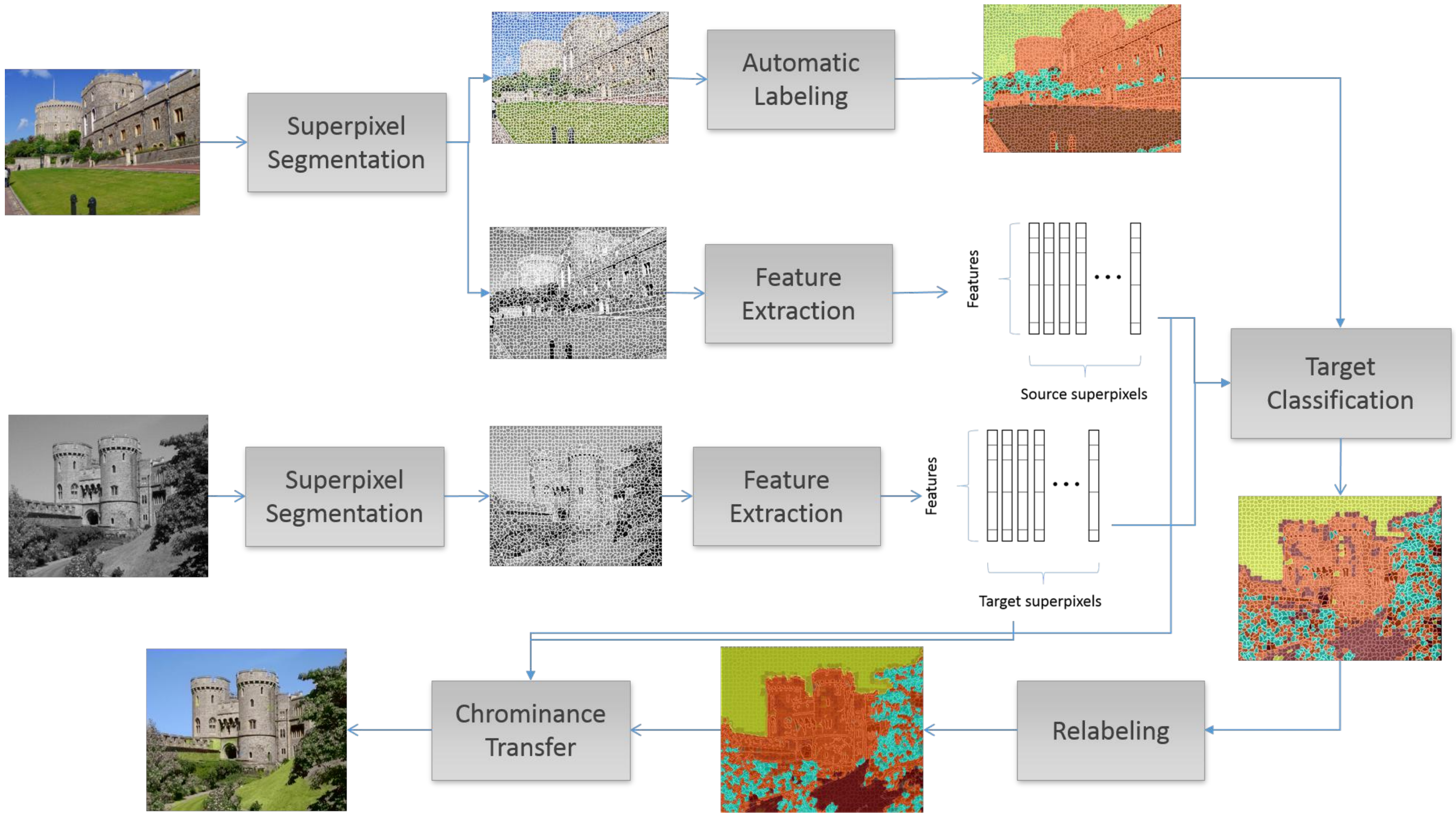


Figure 5 – Abstraction of the method pipeline. Named blocks are the pipeline modules and the images are representations of the intermediate outputs.

The following stages of the method rely on comparing intensity values from both images, therefore their intensity distributions need to be somewhat similar. As stated in (HERTZMANN et al., 2001), the use of *histogram matching* (GONZALEZ; WOODS, 2012) to match the target image histogram to the source causes undesirable effects over the target. Therefore, the proposed method utilizes the simple *linear mapping* from (HERTZMANN et al., 2001) which manipulates the intensity levels of the target image so that mean and variances of its distributions match the source.

After going through this preconditioning stage, the images have comparable intensities, which allows for consistent comparisons in the following stages.

3.1.2 Superpixel Segmentation

Superpixel segmentation is a technique that oversegments an image into non-overlapping structures called *superpixels*. These superpixels are groups of contiguous image pixels that share similar characteristics, which in our application are intensity values. The idea of superpixel segmentation was introduced in (REN; MALIK, 2003) as a preprocessing step for image segmentation in a split-and-merge framework.

Due to its characteristics, the use of superpixel segmentation fits well in the colorization task. First, the superpixels can speed-up the algorithm since it reduces the number of elements for the subsequent algorithm steps, particularly for the computation of the metric space which involves computing pairwise distances between these elements. The exemplar-based algorithms that do not use superpixels rely on random sampling of source pixels to be able to execute in reasonable time. Beyond the performance aspect, there is also a quality improvement to superpixels. Since they group similar pixels into regions, the use of superpixels in itself enforces spatial coherence on the colorization result.

Figure 6 shows the result of performing superpixel segmentation using the *turbopixels* algorithm (LEVINSHTEIN et al., 2009), which is the algorithm featured in the method pipeline as well as in other exemplar-based techniques ((GUPTA et al., 2012), (LI; LAI; ROSIN, 2017), (LI et al., 2017)). This algorithm tries to generate superpixels that preserve the original image structure while maintaining similar size and shape.

After segmenting both source and target images into superpixels, the subsequent steps in the pipeline are carried at superpixel level. One disadvantage of using superpixel segmentation is the necessity of adapting subsequent steps in the pipeline that were originally designed for pixel granularity.

3.1.3 Color Clustering and Source Labeling

The proposed method is based on statistical classification, therefore, each target superpixel, instead of being matched to its closest source superpixel, will be classified into one of the available color classes. Therefore, the method starts by generating this set of classes.

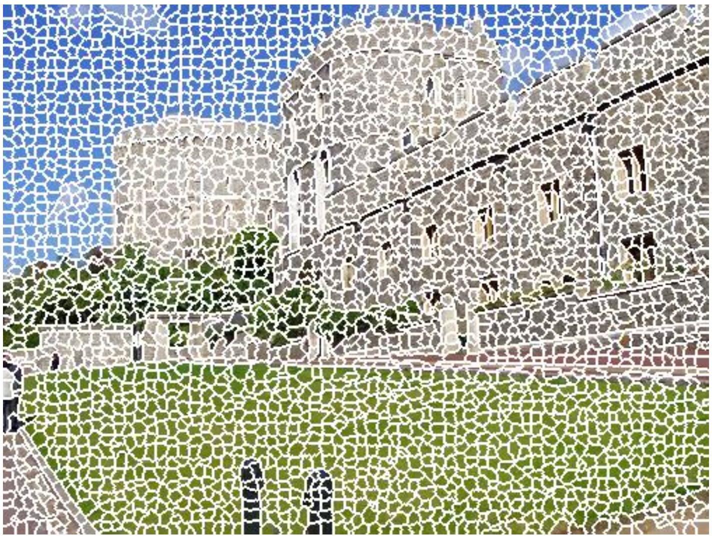


Figure 6 – Example of superpixel segmentation with around 3000 superpixels.

In (IRONY; COHEN-OR; LISCHINSKI, 2005) the authors also used a classification framework in which the classes are defined by requiring the user to provide a partially segmented source image, in which each segment is roughly uniform in color and texture. To reduce the need of user interaction, in this work we opt for a straightforward automatic source label assignment similar to (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008).

Initially, the source image pixels are clustered according to their chrominance values (ab channels of the Lab image) using the k-means algorithm, so that each pixel receives a cluster label. After assigning a label to each **pixel** through clustering, the label of each **superpixel** is defined as the *mode* of the labels within the superpixel. Since the superpixel segmentation has structure awareness, usually the pixels contained within a superpixel share the same label and the mode frequently represents the vast majority if not the full set.

Figure 7 illustrates the label assignment described above. Each pixel in color image on the left is divided between four clusters based on their chrominance generating the middle image. Then the mode is applied in each superpixel to generate the superpixel labels (right side image). The superpixel labels are the output of this stage, as illustrated in Figure 5.

3.1.4 Feature Extraction and Superpixel Descriptors

Exemplar-based methods in general rely on local descriptor similarities to transfer colors, so, the choice of the features that compose the final descriptor plays an important role in the colorization. As the user selects the reference image based on semantic content, ideally the features would be able to translate different semantics into distinct descriptors so the algorithm would be able to match elements by semantic content. Since semantics-based instance segmentation is far from a closed problem, the algorithms rely on simpler local

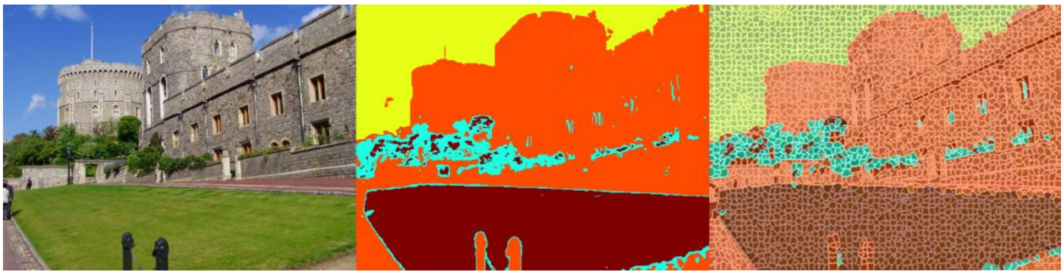


Figure 7 – Color image (left), initially clustered by chrominance at pixel level (middle) and final superpixel labels assigned by pixel label mode (right).

descriptors to compute pixel/superpixel similarities.

The review carried in Chapter 2 indicated that feature extraction for image colorization is not a consensus with authors exploring different combinations of features including: raw pixel values and statistics, dct, dft spectrum amplitude, Gabor filter banks, dense sift/surf, lbp, and others. To use these well known features in the context of colorization, which requires that each pixel receives its own description, features that were originally designed for keypoints (such as sift) should be computed for each pixel instead (*dense*), while the ones that were originally global (e.g. dft), should be computed in *patches* around each pixel. The design of feature extraction at superpixel level requires further considerations.

In designing our feature set from the features presented in the literature, we tried to create a concise set in which each feature complements the others while trying to encompass information of low, medium and high level. The proposed feature set is composed of:

- Intensity: Raw pixel values from the gray images.
- Image Gradient: hog from gray images.
- Gabor filters: Feature volume generated by the responses of the gray images to a filter bank composed of Gabor filters of varying orientations and scales (JAIN; FARROKHNIA, 1991).
- Dense SIFT: The dense version of the scale-invariant feature transform (LOWE, 2004).
- Saliency map: Saliency likelihood of each superpixel in the image (YANG et al., 2013).

Raw pixel values are the most straightforward description of pixels. Since isolated pixels do not carry enough information, intensity-based features are built from statistics (mean, standard deviation) over patches centered on each pixel. Since we work at superpixel level, we adapt the intensity-based features to use the superpixel region itself instead of a generic rectangular window around pixels, such as in (LI et al., 2017). Our

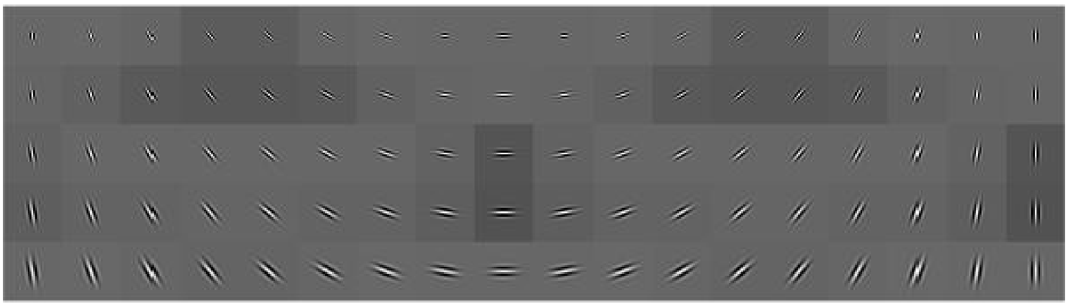


Figure 8 – Real part of the Gabor filter functions with varying wavelenghts (rows) and orientations (columns).

intensity-based superpixel feature is composed of the *mean* and *standard deviation* of the pixel values within the superpixel as well as a *normalized histogram* of these values.

Besides considering the intensity levels, it might be useful to examine the local changes in these intensity values using the image gradient. Since an image is a discrete-valued function, the image gradient might be estimated using finite difference operators along the image coordinates. The gradient-based superpixel feature proposed in this work estimates the *consonance* in the directions of gradient vectors within a superpixel, based on the premise that similar regions should have similar gradient patterns. The algorithm computes the hog for each superpixel and then the ratio of the number of directions in the largest bin from the total. This quantity varies from $1/nBins$, if directions are uniformly distributed to 1, if vectors share the same directions within the superpixel. In order to be rotation invariant, the ratio considers the direction distribution, not the actual direction values.

After analyzing intensity and its local variations, the next feature considers patterns of local repetitions in the images, called *textures* in image processing literature. These patterns might be identified through the images responses to periodical functions, which leads to the use of *spectral* features. Since textures occur in regions of the image, the feature must be able to analyze frequency content throughout the whole image providing responses in a local level. Gabor filters are designed for that specific task. A two-dimensional Gabor function is composed of a Gaussian envelope, which provides the desired locality, modulated by a sinusoidal plane wave of some frequency and orientation, which provides the frequency content analysis. The feature is built from the image responses to a filter bank (Figure 8) composed of Gabor filters of varying scales (frequencies) and orientations. The response generates a volume in which each depth layer contains the response to each filter. The superpixel feature consists of the vector generated by taking the mean of Gabor vectors from pixels within the same superpixel. For more details on space-frequency analysis in image processing, see (HYVÄRINEN; HURRI; HOYER, 2009)

The sift feature is included because of its strong description capability. The sift presents characteristics that makes it view-invariant and these properties might be useful for the

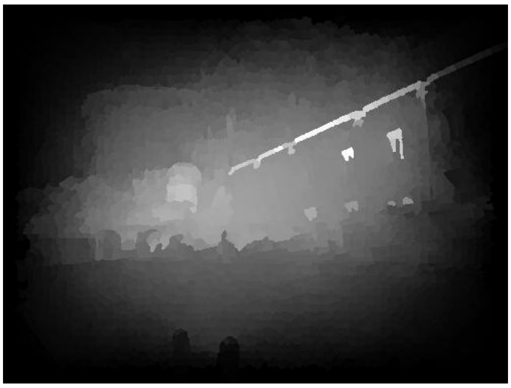


Figure 9 – Saliency map generated for the castle image.

colorization, specially when there are matching objects in the image pair. The algorithm utilizes the mean of the dense descriptor vectors within each superpixel to form the superpixel descriptor, as we did with the Gabor feature

The last feature of the set is based on *saliency maps*. Saliency detection aims at identifying the most important and informative parts of a scene. The premise we follow is that salient foreground objects from the source image are probably good references for foreground elements on the target, so the proposed technique includes these saliency maps with the intent of mimicking a semantic-based high-level feature. Saliency detection was successfully applied as high level feature for colorization in a recent exemplar-based method (LI; LAI; ROSIN, 2017). The saliency map is built by assigning a saliency likelihood to each superpixel in both images, using the algorithm in (YANG et al., 2013) (as shown in Figure 9) which was the same used in the colorization method (LI; LAI; ROSIN, 2017).

After computing all the features, the local descriptor of each superpixel is formed as an heterogeneous vector composed of the concatenation of all these features. The distances between these vectors are used in the following stage to determine target superpixel neighborhoods and then assign classes to them.

3.1.5 Target Classification

The previous section described how to compute local descriptors for each superpixel from both source and target images. The distances between these descriptors (source and target) should be used as a similarity measure for the upcoming color transfer. The proposed method relies on statistical classification as an intermediate step before color transfer, in which first we transfer labels from source to target and then, these labels are used to guide the color assignments.

We decided to avoid direct one-to-one matching because it does not provide robustness to possible *noise* in the feature space. By noise in this context, we mean the presence of descriptors with labels/colors in disagreement with its neighbors in feature space, which is not uncommon, specially due to the multimodality. We expect the method to initially reduce the effect of noise by the use of majority voting based classification.

The classification scheme proposed in this work is performed as a two-step process that initially transfer labels from source to target and then refines the label assignments within the target based on self similarity. The proposed scheme tries to incorporate the most from the available information, including both descriptor distances (3.1.5.1) and distances between color clusters defined in Section 3.1.3.

3.1.5.1 Metric Space Construction

In order to determine the neighborhoods of each target superpixel, first we need to define distances in the feature space. Due to the heterogeneous nature of the descriptors defined in 3.1.4, these distances cannot be calculated as simple euclidean distances between the feature vectors.

In order to have more control over the influence of each feature separately, we decided to compute the distances from each feature separately and then combine these distances with adjustable weights. The pairwise distances matrices are based on *absolute difference* for scalar features, *euclidean distance* for vectors and *match distance* (described in (RUBNER; TOMASI; GUIBAS, 2000)) for the histogram.

One major advantage of computing these distances separately is that we can perform *minmax* normalization over each distance matrix separately and according to the actual computed boundaries (*relative extrema*) instead of according to the *absolute extrema*. Take for instance a feature vector which have N dimensions with each dimension normalized to the $[0, 1]$ range. If we were to normalize the distances between such vectors according to their absolute limits, we would divide each distance by $\sqrt{N}$ because of the set absolute extrema (distance between vectors of all ones and all zeros). But since there is a high probability of the set absolute extrema not being present in the sample set, the distances computed for this feature end up losing their relative influence when compared to a scalar feature which does not go through the normalization. Once the distances are computed separately, we are able to normalize the distances by performing *minmax* using the minimum and maximum sample distances and therefore guarantee that every distance falls in the $[0, 1]$ range. Repeating this procedure for all features allows the algorithm to control feature influence directly through the assigned weights.

The combined pairwise distance matrix is formed by the linear combination of each separate pairwise matrix multiplied by its assign scalar weight. The nearest neighbors of each superpixel can be found by sorting each row of the combined matrix.

3.1.5.2 Class Prediction

The classification of each target superpixel relies on the feature space distances computed in the previous step. To make full use of the aforementioned distances, instead of using the standard majority voting knn, in the classification of superpixels the contribution of each of its neighbors is weighted by the inverse of theirs distance to that superpixel.

Although the knn could also be used in a regression framework instead of classification, the literature advises against regression in colorization due to its averaging effect which causes the final colorization result to be bland as discussed in (CHARPIAT; HOFMANN; SCHÖLKOPF, 2008), (LARSSON; MAIRE; SHAKHNAROVICH, 2016), (ZHANG; ISOLA; EFROS, 2016).

Since the method generates classes automatically by color clustering (Section 3.1.3), if the source image presents an uneven distribution of colors (a particular color is abundant in the scene), the defined classes become unbalanced. *Class imbalance* is a problem for the knn because the feature space is populated with an uneven number of samples and thefore the classification might become biased. We could verify the effect of class imbalance in our early prototypes which generated colorizations dominated by the source abundant colors. To overcome the imbalance issue, the proposed method includes a modified knn similar to the knne shown in (SIERRA et al., 2011). The method consists of the weighted knn described above, but instead of looking for the k nearest neighbors of an instance, it actually considers the k nearest neighbors from each class (with the total neighborhood size of k times the number of classes).

The classification rule followed by the proposed method can be summarized in Equation 3.1.

$$\forall i \in \{1, \ldots, N_{SP}\},\ \hat{y}_i =_{y=1\ldots N_C} \sum_{j=1}^{N_C} \hat{P}(y \mid x_i) C(y \mid j), \tag{3.1}$$

where $\hat{y}_i$ represents the predicted class/label (from the set of N_C available classes defined in 3.1.3) for the target superpixel x_i. The terms $\hat{P}$ and C are respectively the estimated posterior probability of classes given observations and the *misclassification cost*. Instead of assigning labels with highest posterior probability (as in a naive bayesian setting), this equation assigns the label of the class that provides the safest choice. By safest we mean taking into account the costs of *all* the possible wrong outcomes associated with a label if assigning this label (misclassification costs).

Although the classification could be performed by taking into account only the probabilities, the misclassification costs allow for better use of chrominance information available from the source image. Once the classes are generated by clustering (3.1.3), naturally there will be classes closer/more distant to each other in terms of chrominance values. For instance, we want the algorithm to penalize more the misclassification of blue to red or yellow than one of blue to purple. The misclassification cost matrix is built so that each element (u, v) is the cost/penalty of predicting class u when the actual class would be v:

$$C(u \mid v) = |\mu_u - \mu_v|^2.$$

The penalty is basically the euclidean distance between the centroids of each class in chrominance space (μ). Since the distance is a commutative operation, the generated matrix is originally symmetrical. Euclidean distances in the Lab color space mimic our perception of colors differences, so the use of this distance fits the colorization goal. The cost matrix is normalized so that each row/column sums to one in order to avoid introducing bias into Equation 3.1.

The estimated posterior probability of a class given an instance (or *score* of the class) is computed using the the full neighborhoods of each instance. The score of a class consists of the ratio of the sum of weights of the nearest neighbors from this class to the sum of weights of all nearest neighbors (from all classes).

$$\hat{P}(j \mid x) = \frac{\sum_{i \in \eta_j(x)} W(i)}{\sum_{i \in \eta(x)} W(i)}. \tag{3.2}$$

The full neighborhood of each observation ($\eta(x)$) is composed of the union of the k closest superpixels from each class ($\eta_j(x)$). The weights $W(i)$ are given by the inverse of the distances described in 3.1.5.1.

3.1.5.3 Two-Stage Classification

So far we described how to transfer labels from source to target superpixels based on the feature space distances between them. Although the described process is already a classification in itself, the proposed method actually performs a two-stage classification.

As previously discussed in Section 2.2.1, the input pair with similar semantics might be originated from different imaging conditions such as different camera, resolution, lighting, etc, which may cause the descriptors of similar elements to be distinct.

The proposed strategy to handle this *multirepresentability* issue is based on the premise that elements of similar content should generate similar descriptors at least within the same image (self similarity principle (BUGEAU, 2018)). The strategy consists of an initial *inter-image* label assignment stage followed by a *intra-image* label refinement stage and we expect to achieve a higher level of target classification coherence after this refinement.

Both classification stages are based on the prediction scheme described in the past section (3.1.5.2), but in each stage we apply a different set of feature weights for the metric space construction (3.1.5.1). In the first stage (inter-image), the method transfers labels from source to target superpixels. Since this initial mapping is expected to be more difficult due to the presented reasons, the method emphasizes (weight-wise) the sift feature, since we expect this feature to be more robust due its highly descriptive nature. The saliency-based feature is also considered since it can facilitate the mappings for image pairs that contain noticeable elements. Then, the second stage (intra-image) performs a relabeling of each superpixel in a process inspired by the Leave-one-out cross-validation scheme. The new label of each target superpixel is defined by the prediction method in 3.1.5.2, but instead of using the source labels as reference, this step uses the target labels from the first stage as the reference. In other words, the neighborhoods of each target superpixel from Equation 3.2 are generated by the target superpixels, but of course not counting the instance in its own relabeling. In this stage, the method emphasizes the texture-based features, since they can reliably identify similar regions within the image.

The output of the “Target Classification” block in Figure 5 illustrates the result of the classification process. The algorithm transfers the labels from source superpixels (represented by the four colors) to the target superpixels.

3.1.6 Edge-Aware Relabeling

In the previous section, we described how the proposed method performs the *local* prediction of classes for each superpixel. Even though the method has a robust classification scheme, the literature review indicated that, as robust as the matching/classification technique might be, errors at local level are bound to happen, due to noise in feature space and multimodality.

Most exemplar-based algorithms rely on the post-processing of the colorized image to correct the errors from the local color assignment and to enforce spatial coherence. The post-processing techniques usually include some filtering or regularization of the chrominance channels, which can cause loss in image contrast, specially around boundaries.

Instead of post-processing the already colorized image, we propose a simple yet effective heuristic that performs a relabeling of the locally assigned labels by taking into account the location of edges on the original grayscale image. The premise is that texture-rich regions have more potential to be classified correctly by the local predictions while regions that lack texture possess less distinctive descriptions and therefore are more prone to wrong local labeling. The relabeling technique aims to improve the robustness on the smooth regions by aggregating superpixels that are confined within the same image boundaries, changing the classification from a local level to a cluster level. These clusters are formed by applying the *Canny* edge detector to the image and uniting superpixels based on the absence of edges between them.

Edge-Aware Clustering

The algorithm first determines each superpixel spatial neighbors (not to be confused with feature space neighbors) by applying morphological operations to the superpixels masks. Each superpixel region is dilated and the overlapping regions determine the neighborhood.

The neighbors lists along with the Canny edge binary image and the superpixels centroids coordinates are used to create the clusters by performing a connected component labeling process at superpixel scale through a simple graph traversal algorithm, considering each superpixel as a vertex and the lists of neighbors as connections. The pseudocode of the clustering algorithm is shown in Algorithm 1.

The algorithm starts with all the superpixels as free elements. Then, it creates a new cluster by picking the first element in the free list and performing an expansion (*Expand-*

Algoritmo 1: Edge-Aware Clustering.

```
inputs : cannyEdges: the Canny edge binary image
         centrsSPs: The superpixels centroids coordinates
         NeighborsLists: the lists of superpixels spatial neighbors
output: clusters: the edge-aware clusters
Function EdgeAwareClustering(cannyEdges, centrsSPs, NeighborsLists):
    freeSPs ← {1, ..., centrsSPs}
    clustersList ← $\emptyset$
    $i \leftarrow 1$
    while freeSPs is not empty do
        clustersList($i$) ← freeSPs(1)
        freeSPs ← freeSPs \ freeSPs(1)

        ExpandCluster(clustersList($i$), NeighborsLists, freeSPs, centrsSPs, cannyEdges)
        $i = i + 1$
    end

Function ExpandCluster(cluster, NeighborsLists, freeSPs, centrsSPs, cannyEdges):
    $i = 1$
    for every $SP_i$ in cluster do
        neighborhood ← NeighborsLists($SP_i$) ∩ freeSPs
        clusterNeighbors ← checkPathEdges(neighborhood, centrsSPs, cannyMask)

        cluster ← cluster ∪ clusterNeighbors
        freeSPs ← freeSPs \ clusterNeighbors
    end
```

Cluster function) from this first element. The expansion is performed by continuously adding to the cluster the neighbor superpixels of each of its elements that do not have an edge in between. The function *checkPathEdges* is responsible for evaluating if two superpixels have an edge element in between them. The expansion process of a cluster ends when there are no more superpixels that can be reached without going through an edge. Then, the whole cycle of creating and expanding a new cluster repeats until there are no more free superpixels, in other words, superpixels without a cluster index.

It is possible (and common in texture-rich regions) that clusters are formed by single superpixels, in which case the original label stays unchanged. The result of this spatial clustering technique is shown in Figure 10. It is noticeable that texture-rich regions form many small clusters while smooth regions group up which is the desired outcome.

The function *checkPathEdges* is as a simple heuristic that verifies if there are any active pixels in the edge mask within the rectangle formed by centroids of the two superpixels being evaluated. The function purposefully overestimates the presence of edges because regions merged incorrectly might generate rough colorization errors, while if regions that could be merged are kept separated, the result at most remains the same.

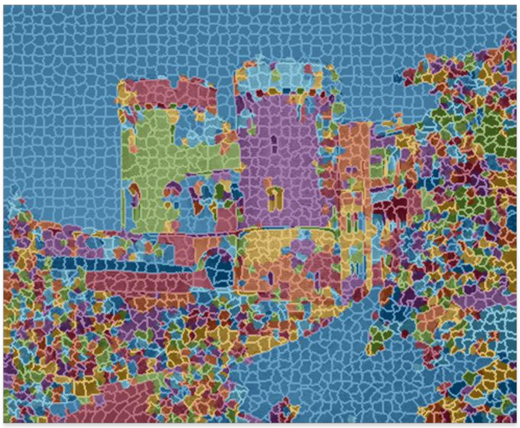


Figure 10 – Edge-Aware clustering of target image. Texture-rich (trees area) regions create many clusters while smooth regions (sky) tend to form large bundles. (Repeated colors that are not contiguous are not from the same cluster.)

Once the edge-aware clusters are defined, the relabeling process consists of simply assigning to all the superpixels within a cluster the label mode of that cluster. The Figure 5 illustrates how the relabeling is able to change the original label assignment of superpixels.

3.1.7 Chrominance Transfer

The final step of the proposed colorization method consists in transferring colors from source to target based on the target final labels and again on their local descriptors. The colors are transferred only between source superpixels that share the same label as the target.

For each target superpixel, the algorithm initially selects the source superpixels that have the same label as the target. Then, from this subset, the algorithm chooses the source elements that are closest to the target in feature space (nearest neighbors). From these nearest neighbors, the algorithm extracts the *median* of their chrominance, which is composed of the median values of each chrominance channel (ab from Lab) within this nearest neighbors set. Again the method relies on multiple neighbors to enhance robustness.

The median chrominance values are transferred only to the centroid of their corresponding target superpixels (as the *micro-scribbles* in (IRONY; COHEN-OR; LISCHINSKI, 2005) and (GUPTA et al., 2012)). The colors are then propagated from the micro-scribbles to the whole image using Levin's algorithm (LEVIN; LISCHINSKI; WEISS, 2004), which was described in the review 2.1.1. The use of this scribble-based technique also enforces spatial coherence.

The color propagation yields the fully colorized output image, illustrated at the pipeline end in Figure 5.

3.2 DESIGN AND IMPLEMENTATION CONSIDERATIONS

During the design and implementation cycle, different approaches were explored until the proposed method reached the configuration described in this chapter. Although the key design decisions will be further explored in Chapter 4, there are other aspects worth mentioning.

Regarding the feature set, different features from the literature were tested. The use of *window-based* features (i.e. features that are functions of a rectangular windows around the pixel of interest) revealed that they deteriorate the matches around image edges, generating halos, which gets worse as the window size increases. Although this haloing effect gets attenuated by the use of superpixels, a local texture feature that uses a window of decreasing weights (such as the Gabor filter) shown to be more robust.

Still on feature set design, both the feature choice and weights have an impact over the final result. Initially, we wanted to evaluate the impact of feature choices in the arrangement of the feature space using data complexity measures such as overlapping, but this task has proven to be very difficult.

About the superpixel segmentation, it is indeed an advantageous approach, both in terms of cost, reducing the search space of classification in orders of magnitude without the need of random sampling, and also providing robustness to the colorization by enforcing spatial coherence.

4 EXPERIMENTS AND EVALUATION

After presenting the proposed model and its characteristics in Chapter 3 we move on to describe the experimental part of this research.

The experiments in this chapter are intended to verify the design decisions and the claims made in the previous chapter and also compare the proposed method with state of the art algorithms. More specifically, we want to verify if the proposed method is in fact able to handle a broader set of input image pairs when compared to the literature exemplar-based methods.

The reported results from exemplar-based methods in the literature feature almost exclusively simple scenery composed mostly of landscapes. By submitting the literature algorithms to other kinds of images, limitations in terms of supported scenes become apparent. The images that are more difficult for the exemplar-based algorithms are images where regions of similar local descriptions appear in different regions of the reference image, which might have different semantics and therefore different colors (IRONY; COHEN-OR; LISCHINSKI, 2005). This characteristic might be present in almost any image that is not a simple landscape, such as: faces, object-centered images, colorful images, cluttered scenes, among others.

We expect the proposed algorithm to generate better results in scenes of higher complexity because of the the designed decisions described in the previous chapter, especially:

- The use of multiple neighbor classification instead of matching that enhances robustness to feature/descriptor space noise.
- The two-pass classification scheme that allows for a more systematic use of the feature set.
- The edge-aware clustering heuristic that combines local decisions into region-based decisions.

4.1 EXPERIMENTAL PROTOCOL AND EVALUATION

The proposed method prototype and all its variants were implemented as MATLAB scripts, the source code is available at Github [1]. In the comparisons carried in Section 4.3, the results of the baseline methods were generated using the authors' own MATLAB implementations (also available online) with their respective default parameters. Execution time was not considered in the comparisons because the published code for the baseline methods were not optimized for time efficiency and therefore the comparison would be unfair.

[1] https://github.com/saulo-p/ImageColorization

The main result evaluation method used in this document is *visual inspection*. Although visual criteria are subjective, evaluating image colorizations by comparing to ground truth is not a valid option, first because the color ground-truth might not be available, and also because it does not fit the goal of colorization as discussed in Section 1.1. Through visual inspection we assess the visual quality of the output images and also evaluate local details of the result. Moreover, for the evaluation of exemplar-based methods, it is important to assess how much of the source image "appearance" is actually transferred to the target instead of analyzing the output by itself and this can be best achieved through visual assessment.

Each image in this chapter is embedded in the document with its original resolution, so, for detailed inspection, the reader can zoom-in until desired size is achieved.

4.2 METHOD DESIGN DECISIONS

In this section we evaluate the impact of key design decisions on the proposed method by visually comparing partial results generated by method variants. The partial result images are formed by assigning each superpixel entire region with a single color from its closest superpixel in feature space.

As the entire superpixel receives a single color, without any smoothing on region transitions, the partial result images present block artifacts. However, the colorization quality of the partial result images is not important for the analysis in this section.

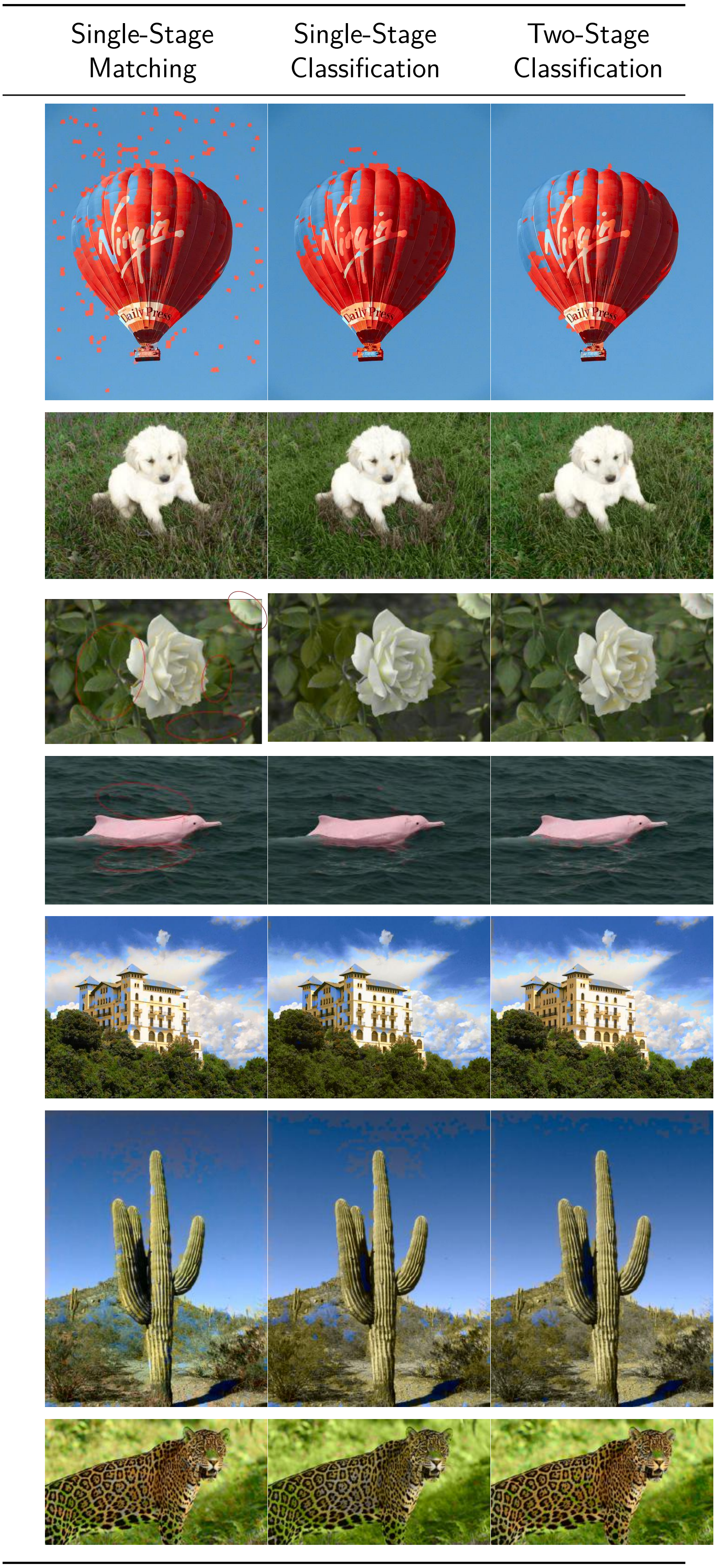
4.2.1 Single-Stage *vs.* Two-Stage Classification

As mentioned in Section 3.1.5.3, the proposed method performs the feature space connections in two steps. Although this choice has a valid reasoning, it is important to verify its practical impact over our classification and colorization scheme.

We observed in the experiments that the two-stage classification have a positive impact over the colorization results in most cases. Figure 11 shows examples in which the two-stage method contributes for a more consistent partial result. We compare single-stage variants using matching and classification with the proposed two-stage classification.

The overall impression from Figure 11 is that the matching variant generates the worst results, as we would expect. The presence of noise in the feature space and/or similarities between descriptors penalizes the approach that assigns colors based on a single neighbor. This behavior is clearly illustrated in the first row of the figure, that contains the balloon image which is a hard colorization image due to the absence of textures that makes the local descriptors of both sky and balloon similar. The consequence is the matching-based approach making wrong color assignments of superpixels scattered over the image. This phenomenon also happen in the second row (zoom in for details).

Figure 11 – Partial results comparing variants of the proposed method.



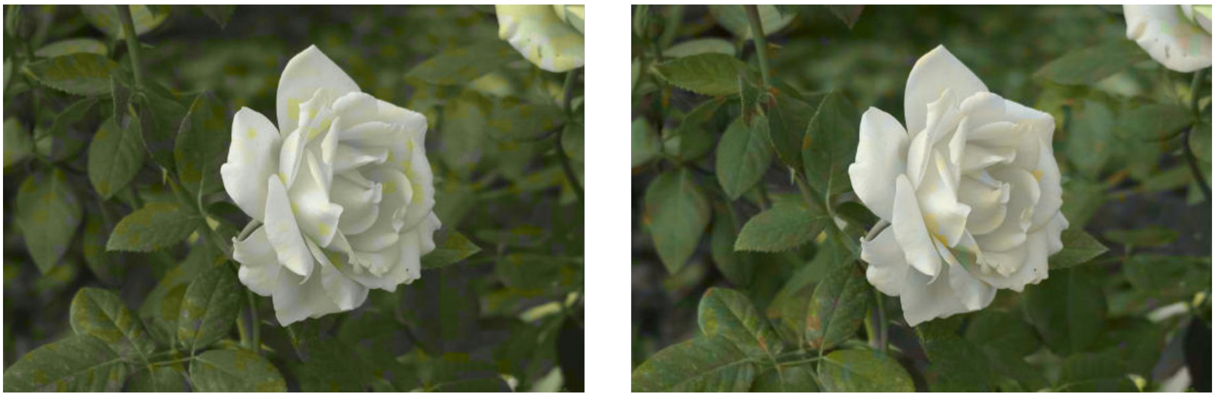


Figure 12 – Left and right are respectively the first and second stages of the proposed method classification.

The single-stage classification variant does not generate as many scattered errors as the matching due to the knn being more robust to the feature space noise. But the two-stage approach can improve the results even further.

The flower image (row 3) is a good example of a successful use of the proposed two-stage method. Apart from the other errors, we want to draw attention to the flower on the top right corner. None of the single-stage variants in Figure 11 is able to correctly assign colors to the corner flower despite correctly assigning to the one in the center. In the first stage of the two (Figure 12), which focus on surf and the saliency-based feature, the corner flower is not correctly classified, in the second stage, the algorithm refines the initial classification based mostly on textures, therefore, due to the texture similarities between the flowers within the same image, it is able to relabel the corner flower region to its correct class. The method achieves the improvement without resorting on a different feature set, changing only the weights in each stage.

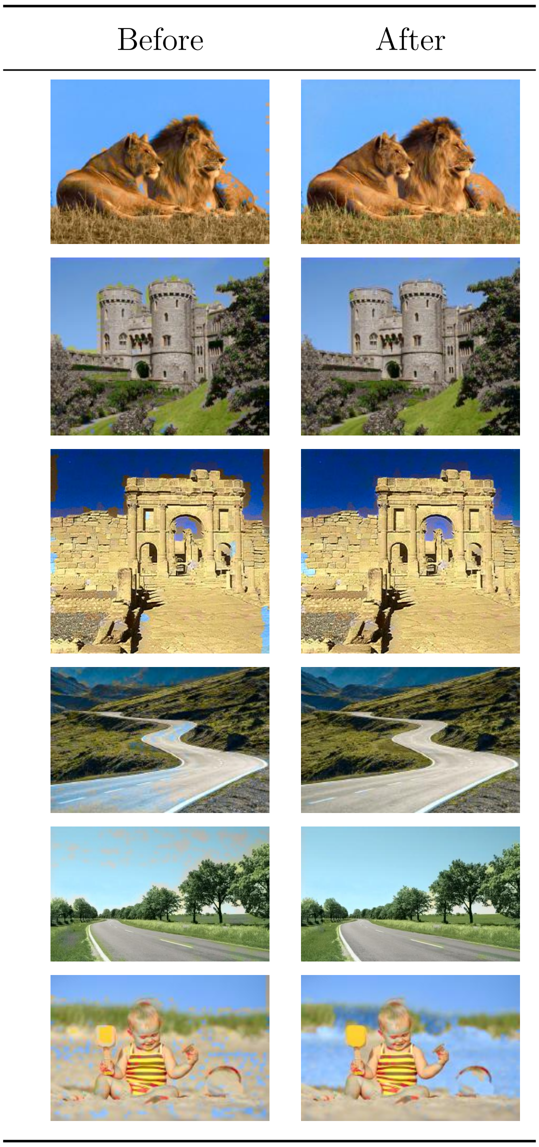
Other images from Figure 11 present other minor improvements. These improvements are usually in the direction of causing the result image regions with similar content to present similar colors.

4.2.2 Local decision x Edge-Aware

Another aspect of the proposed method that deserves evaluation is the edge-aware label refinement described in Section 3.1.6. This particular step is expected to enhance the colorization on smooth regions of the image, where there are no much texture information to serve as clue, while not changing assignments in texture rich regions, where the local descriptors are expected to perform well already. The images in Figure 13 show the color labels of the superpixels before and after the edge-aware relabeling.

The three first rows of Figure 13 show a recurrent error in classification that happens around image boundaries. Even though the superpixel segmentation mostly respect the original structure of the image, the boundary regions in natural images are usually regions

Figure 13 – Edge-Aware Relabeling results.



of smooth transitions of intensity levels. The transitional pixel values cause the superpixel classification around these areas to be more chaotic. The edge-aware relabeling leverages on the fact that most of the superpixels are in plane (not transitional) regions and therefore the majority voting should be able to rectify the boundaries misclassifications.

The image in the fourth row illustrates another advantage of this approach. The local descriptors of sky and road on the source image are similar, and therefore the classification of the road in the target image oscillates between these classes. Due to the clear edges on the road, the algorithm is able to group the whole portion in a single cluster and then assign its majority class (road color), generating the more coherent result. In the fifth row, a similar relabeling effect happens in which incorrect gray colors get removed both on the sky and grass (bottom left). The sky regions between the trees were not connected to the main sky cluster and ended up with a different color. These examples indicate that the edge-aware approach is a possible workaround for the multimodality issue that does not require global optimization (which can lead to bland colorizations).

Although this relabeling step clearly enhances the results in most cases, the strategy can also backfire. The figure in the last row shows an instance in which the relabeling is not a good approach. Due to the lack of closed boundaries in the blurred regions of the image (background), the edge-aware clusters formed are not faithful to the original image structure which causes the most abundant class (sky color) to take over other image regions that were not originally connected.

4.3 PROPOSED METHOD *VS* EXEMPLAR-BASED

After evaluating the key modules of the proposed method in isolation, in this section we move on to assess the technique's final results by comparing it with baseline methods from the literature.

We chose to evaluate the proposed method in comparison with the methods (GUPTA et al., 2012) and (PIERRE et al., 2016) (which is a MATLAB tool that implements (PIERRE et al., 2015)) because they are both fairly recent, report interesting results and their authors made the source codes available online.

We divide the test images into different sets according to the complexity of their depicted scenes.

4.3.1 Simple scenery

The set of images used for evaluation in the baseline work (GUPTA et al., 2012) is composed mainly of simple images, which is a good starting point for the analysis. We consider these images to be simple because texture differences are well correlated with color differences, therefore the local descriptors should be able to discriminate colors fairly well.

The images in Figures 14 and 15 indicate the proposed method results are at least as good as those of the baseline in this class of images. The exemplar-based algorithms generate acceptable to good colorizations for both subsets, with the method (PIERRE et al., 2016) having some issues with the second subset.

First, in the landscape subset, the rich textures and absence of salient objects can cause minor mistakes in color assignment to pass unnoticed by the observer. However, careful visual inspection (zooming in the images) reveals mistakes from the methods.

Since the exemplar-based methods do not possess any level of *scene understanding*, these algorithms usually make semantic mistakes such as using different colors for the same scene elements (leaves of the same tree in rows 1, 2 and 5 of Figure 14) or the reflections of water that are not consistent with the scene (as in rows 2 and 5 of the same figure).

For the superpixel based approaches (proposed method and baseline (GUPTA et al., 2012)) small elements and fine image details can cause flaws in the colorization. Whenever

Figure 14 – Simple images (nature landscapes).

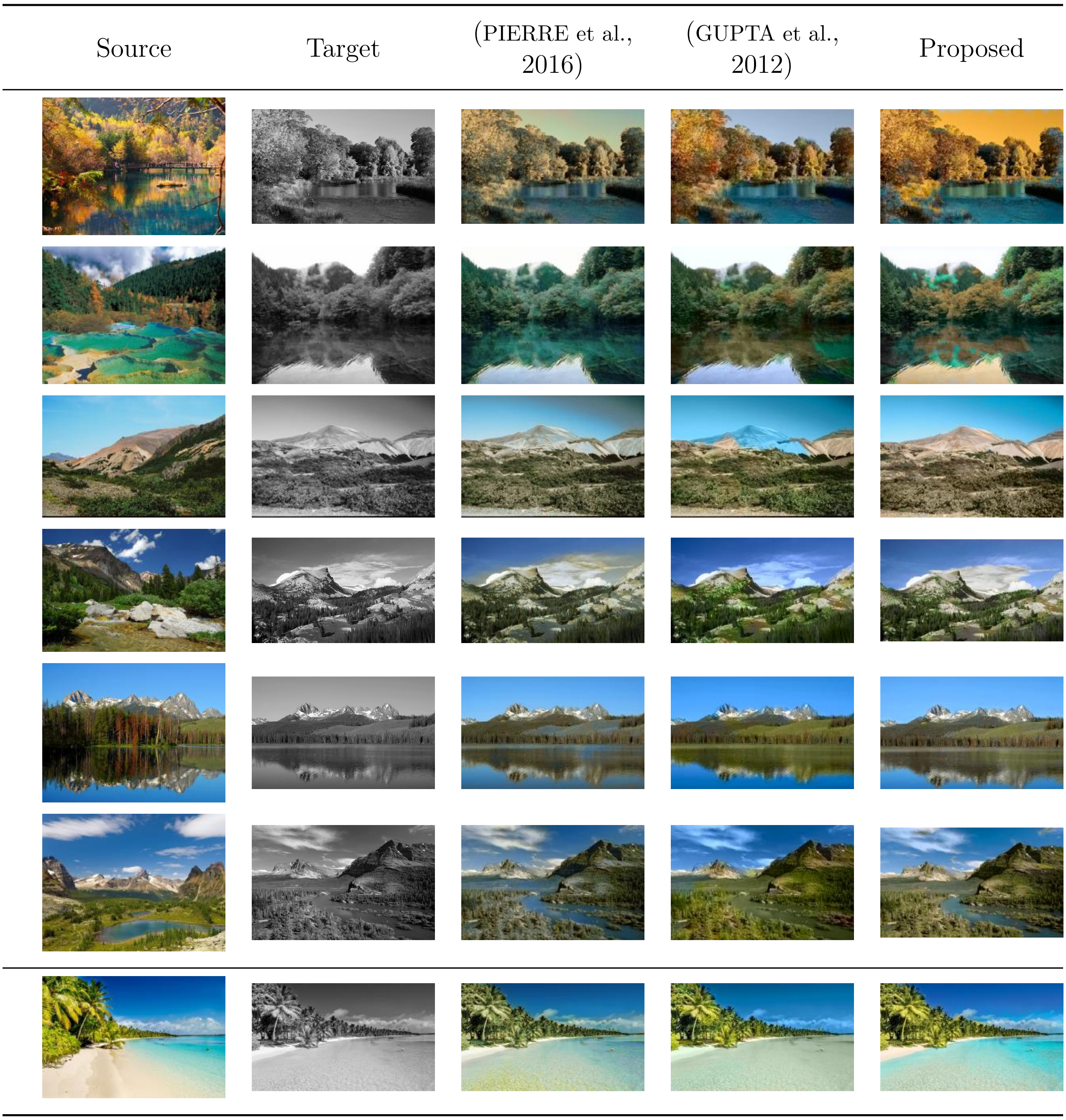
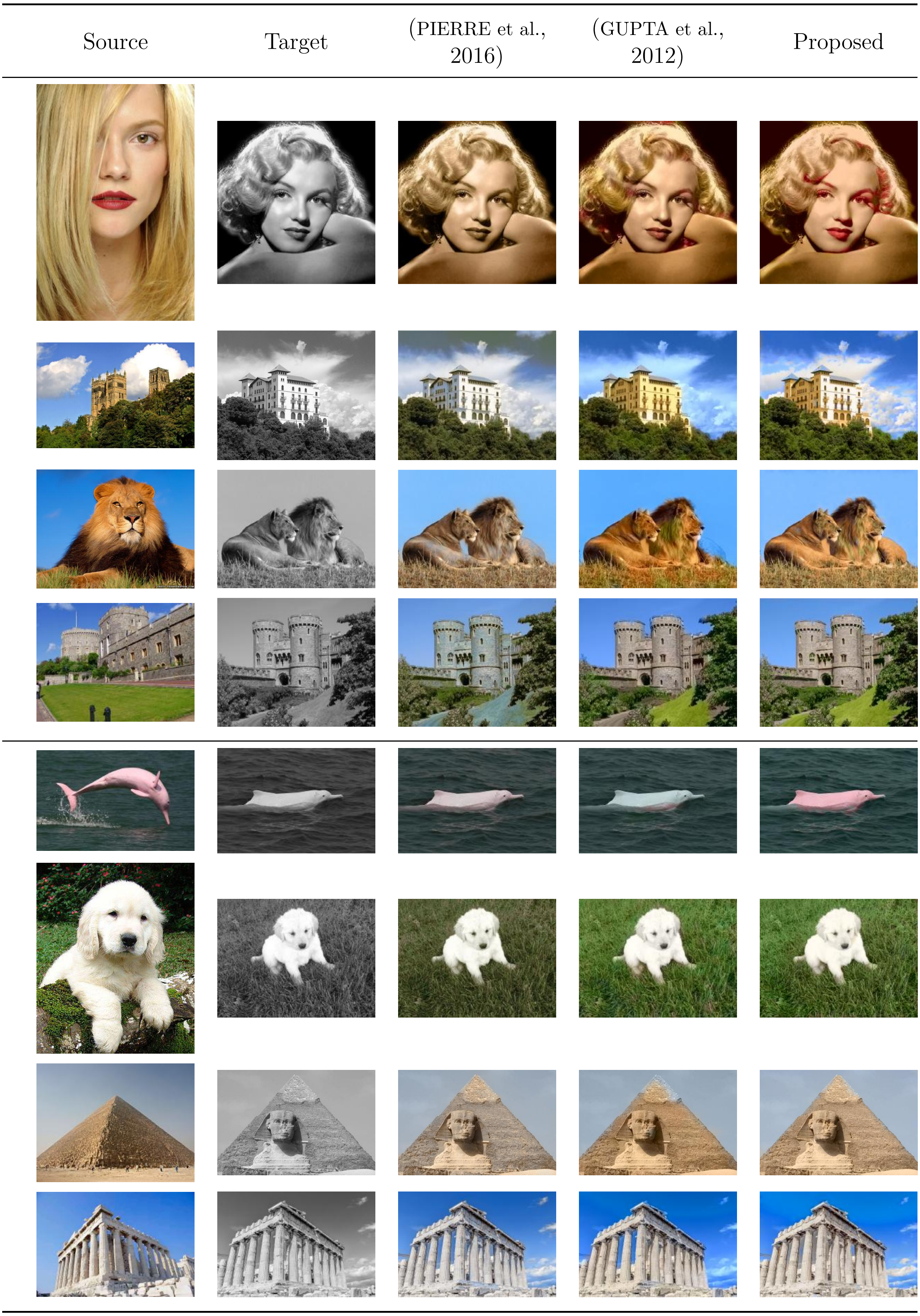


Figure 15 – Simple images (misc).



the superpixel scale is coarser than some image details or fine structures, these regions might be “miscolorized” since the colors in these techniques are assigned at superpixel scale. The small trees in the fourth, fifth and sixth rows from Figure 14) illustrate how the bleeding artifacts arise around the finely detailed areas.

The proposed algorithm generated better results compared to the baseline when required to discriminate colors at regions of smooth transitions as shown in the third row from Figure 14 (top of the mountain) and also in the third row from Figure 15 (at the lion’s back). This phenomenon also occurred at the cloud regions in images 4 from Figure 14 and 2 from Figure 15. Not only the colors around boundaries are more often predicted correctly, the boundaries themselves are more well defined (as in the fourth row from Figure 15, on the top of the castle and around the tree boundaries).

The beach image (last row) is another example of how the proposed technique can generate better results. The baseline (PIERRE et al., 2016) have issues to discriminate between textures of the seashore and trees, while the method (GUPTA et al., 2012) again misses the smooth transition and merged sand to water and clouds to sky. The proposed algorithm, despite a few local green spots in the clouds, generated the best result correctly transferring color between corresponding regions while also exploring the different color tones for the water.

The results might also differ considerably between the algorithms (or even between executions of the same algorithm with different parameters) while still presenting similar levels of plausibility as in the first row of Figure 14 with the different sky colors or in Figure 15 second row with the building colors.

The results from the proposed method and the baseline (GUPTA et al., 2012) are in general better than (PIERRE et al., 2016). The partial results from method (PIERRE et al., 2016) reveal that in many instances (such as the mountain top, the lion’s skin and the castle walls) the initial colorization consists of pixels that have correct color assignments mixed with pixels with mistaken assignments in same regions. The use of superpixels by the proposed method and (GUPTA et al., 2012) enforces spatial coherence and instead of generating a colorization that looks like a blending of different colors, utilizes a single color for each superpixel region.

4.3.2 Complex scenery

Beyond local mistakes in some instances, the proposed method and the baseline (GUPTA et al., 2012) seem to be able to achieve fairly successful colorizations of the simple scenes. In this section, though, we aim to expose the exemplar-based algorithms to a set of more complex scenes.

The Figure 16 displays some selected complex source/target pairs and the respective results from each one of the methods. The fact that the result images are usually worse if compared to the results of the previous section indicates that these indeed are more complex images.

First, in the baby image (second row), the presence of background blur that causes textures to be lost summed with different background colors confuses the color assignments to the point the baseline (GUPTA et al., 2012) mainly uses a single color for the entire image. The proposed method and the baseline (PIERRE et al., 2016) are able to partially identify the image regions but the overall results are still very poor. Images with too many colors, as the fruits in row 7 are also very difficult instances since the many color options available for each pixel/superpixel makes so that the chances of correctly assigning colors naturally decline.

Moving on to the cases of relative success, the image of the climber (first row) exhibits color imbalance. The shirt in the source image has a distinct color that covers only a small portion of the image. Since both the proposed method and (PIERRE et al., 2016) are able to correctly identify the shirt region, the local description of the region must be distinctive enough. Therefore, the method (GUPTA et al., 2012) not assigning the correct shirt color might be due to its post-processing that distrusts the color assignments of small regions surrounded by a different color. The edge-aware relabeling scheme of the proposed method on the other hand maintains small regions colors as long as they have clear boundaries and therefore can be considered an improvement over the baseline in this sense.

The image of the beer in the fourth row is another instance of small region color transfers. The proposed method was the only that transferred the blue color from the chessboard pattern from source to target. The use of color clustering for source labeling (see 3.1.3) combined with the knne classification guarantees that even colors from small regions in the source will be considered during the target label assignment and therefore can be present in the output. On the other hand, small blue spots are present in random areas in the output image, which suggest that our approach might benefit from some sort of tuning to achieve a better balance, but this was not addressed in this research.

The cars images (rows 5 and 6) are examples that put emphasis on foreground elements. The candidate selection from (PIERRE et al., 2016) is not able to correctly transfer the colors of the image elements in none of the two images, which might be due to the feature set employed. While at the fifth row both the proposed method and (GUPTA et al., 2012) generate decent colorizations, at the sixth row, the background blur on both source and target images makes the matching/classification of superpixels more challenging, which causes the method (GUPTA et al., 2012) to miscolorize the car while the proposed method assigns the car colors to some background superpixels. The proposed method correctly transferred the colors between the major elements at the cost of a few local errors.

Another instance of background blur occur in the butterfly image (fourth row), again

Figure 16 – Complex scenery.

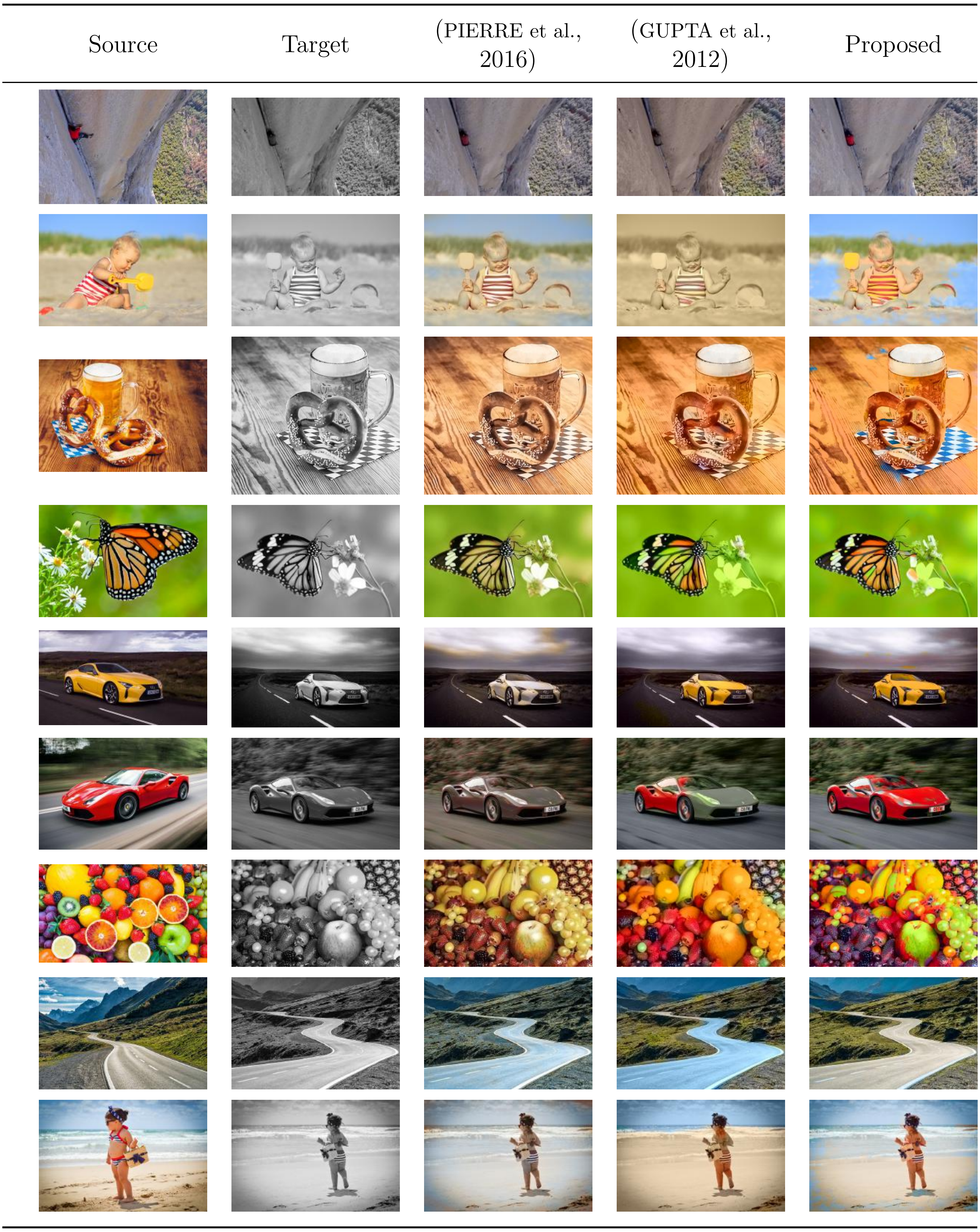
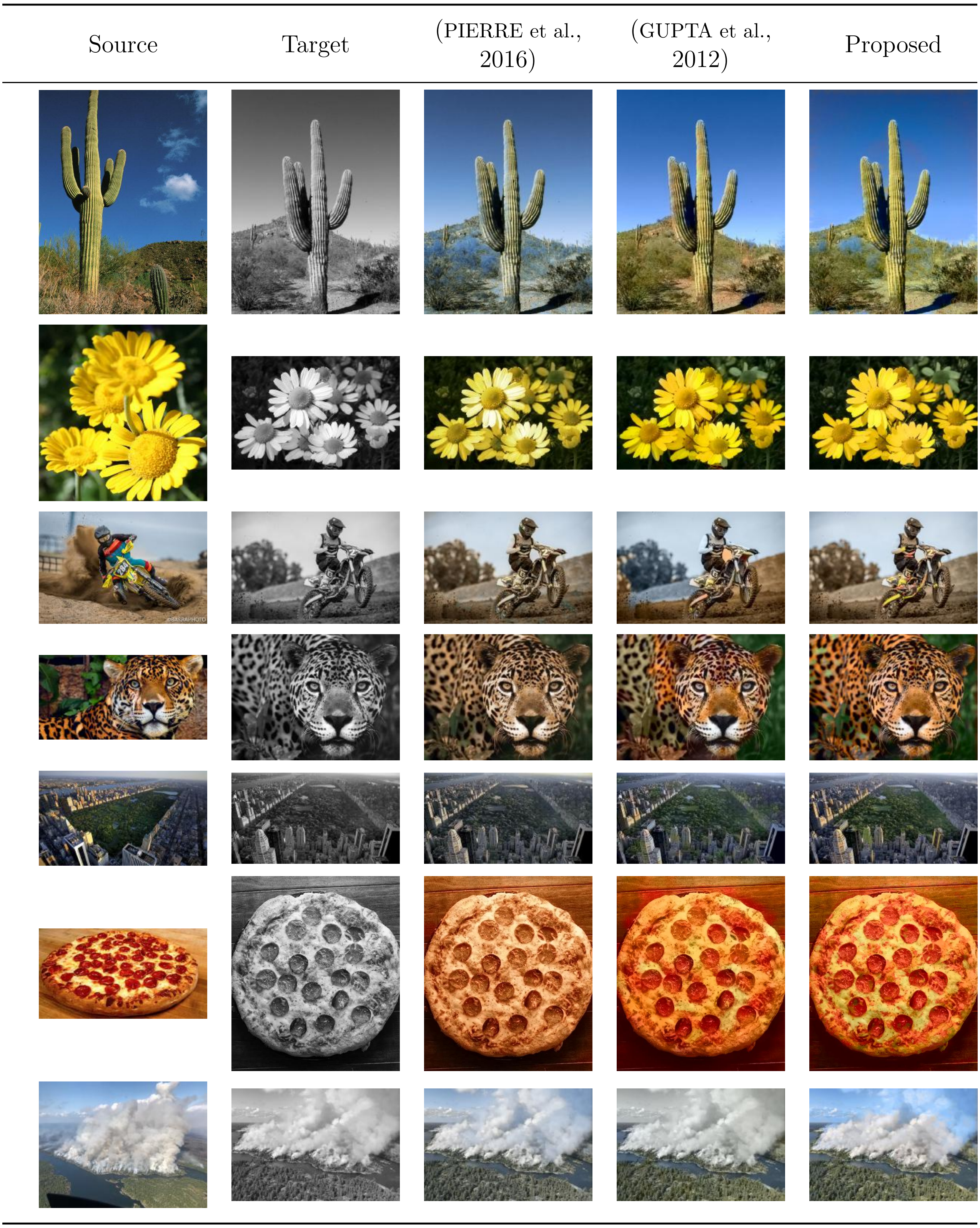


Figure 17 – Complex scenery (continued).



causing complications in discriminating between the blurred background and smooth parts of the foreground (butterfly wings). The result from (PIERRE et al., 2016) presents the color blending effect discussed in the previous section, while the method (GUPTA et al., 2012) assigns the background colors to the majority of the butterfly wing. The proposed method was able to transfer the wings color mostly correctly at the cost of some inconsistency with the flower colors.

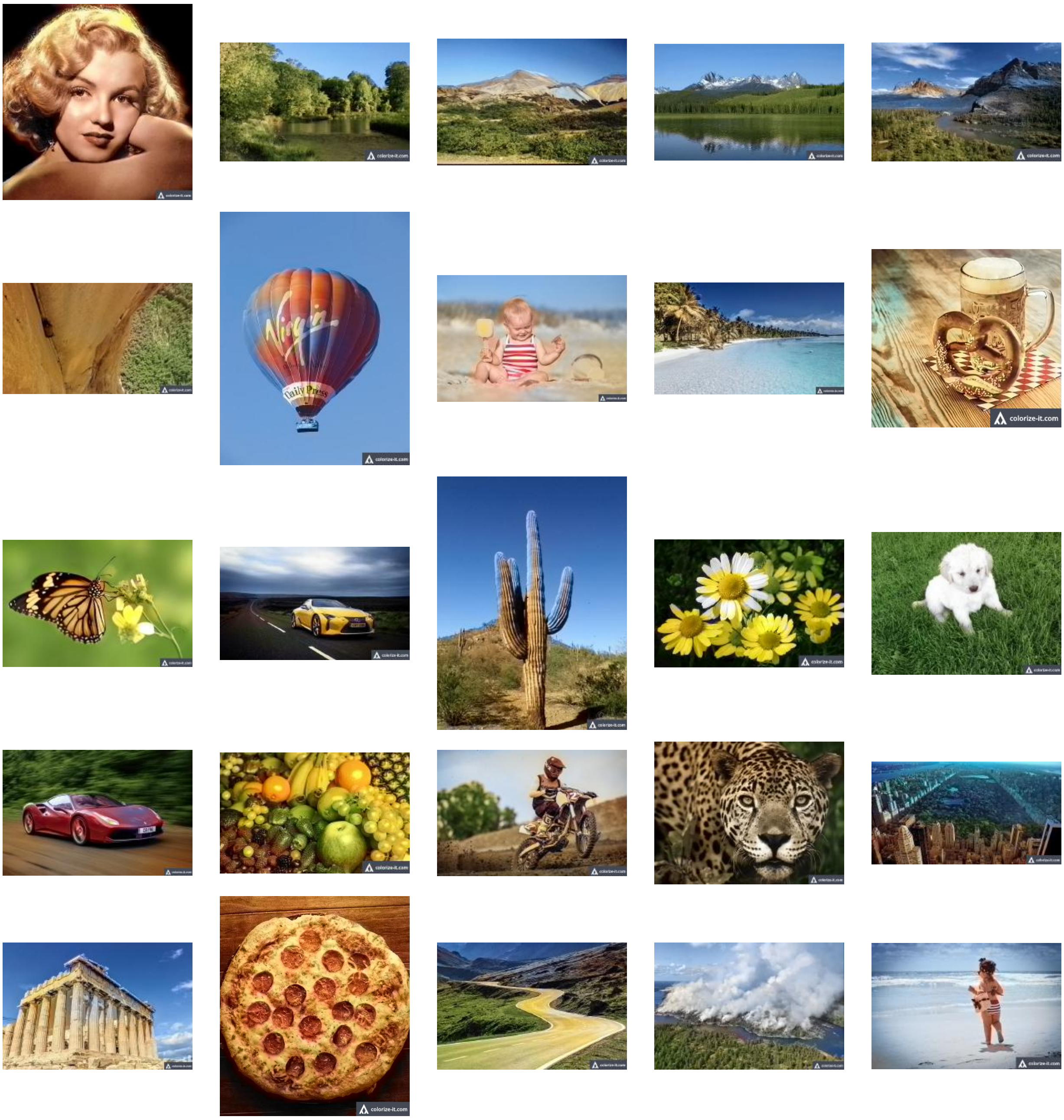
As previously discussed in Section 4.2.2, the road image in the seventh row illustrates how the edge-aware relabeling serves as a workaround for the multimodality problem. The proposed method was able to correctly transfer colors between corresponding image elements of source and target. The robust classification of the proposed method is able to correctly classify (at local level) the majority of superpixels within the road region. Then, the edge-aware relabeling rectifies the local mistakes by incorporating labels from the whole road portion. By improving the spatial coherence at the road area, the method also improves the result in a semantic level, assigning the correct colors to the road while respecting the original image boundaries.

The toddler walking on the beach (row 8) is another complex instance that includes small image details and similar descriptors. The baseline (GUPTA et al., 2012) again showed issues with smooth transitions from sand to water (same as shown in Section 4.3.1) and assigned the sand color for the majority of the water region. The proposed method on the other hand was mostly affected by the intensity changes in the halo that surrounds the image, causing wrong color assignments in the image borders. Other than that, the proposed method result is better than the baseline results, correctly transferring the colors between the background regions while also being able correctly assign colors to the small details of the bikini and purse.

In the second part of Figure 16 we present more results on complex scenes. In the cactus image, the baseline (GUPTA et al., 2012) actually presented the best results while the proposed method made mistakes with the blue color around the image. In the motocross image, the proposed technique generated the overall best result, specially at the motorbike details, but none of the methods was able to transfer the shirt colors from the source. In the park image, the proposed technique was slightly better at discerning between the colors of trees and buildings and was considerably better at transferring the yellow tones to the buildings facing the sun. The smoke image again showcases how the proposed method is better at dealing with smooth transitions, as was the case with the clouds.

Though not always realistic, the proposed method generates less failed colorization and most of the errors are local (not generalized), the impression from Figure 16 is that the proposed method generates results of better visual quality compared to the baseline exemplar-based methods.

Figure 18 – Learning-based colorization results.



4.4 COMPLEX SCENES *VS* LEARNING-BASED

To better situate the proposed method within the bigger picture of colorization algorithms, in this section we present the results of a recent learning-based method. The purpose of this section is to assess whether a recent learning-based method also has difficulties with complex images as the exemplar-based ones, and how its results compare to the results presented by the proposed technique. The output images in this section were generated using as input simple and complex images from previous sections colorized by Zhang's algorithm (ZHANG; ISOLA; EFROS, 2016). The results are presented in Figure 18.

First, since the learning-based methods do not rely on user input, there is no user control over the output images and therefore no guarantee of similarity between the colorized images from the previous section and the ones in this section. Furthermore, the absence of a context reference for the target gray image can make it hard for this algorithm to identify the correct semantic of specific image elements. For example, the climber shirt in the second row gets "ignored" within the mountain color and the paved roads from both the red car image (third row) and the mountain image (last row) which receives the colors of dirt roads.

Regarding the visual quality of the output images, the learning-based method seems to also have difficulties with the complex images. The presence of blur makes it difficult to identify details within the blurred area and therefore the algorithm assigns a single color to the whole area disregarding the different elements within it, such as the park area (third row) and the toddler on the beach (fourth row). Also, separated elements of same semantics not necessarily share similar colors as observed under the biker left arm (third row) and between the columns of the Parthenon (fourth row) which should share the sky colors. Results might also lack in spatial coherence as the blue stains in both the red car (fourth row) and red balloon (second row). Finally, the colorizations in Figure 18 show problems around object boundaries, with the colors bleeding through the object boundaries to neighbor regions as in the top of all the mountains in the first row, the tree tops in the second row, the cactus in the third row and others.

4.5 DISCUSSION

The results and analysis presented in this chapter indicate that the proposed method is a viable option for colorization which generates plausible results for simple image pairs while being able to target image pairs that present complex traits with moderate success.

Our technique shows improvements over previous exemplar-based methods in terms of visual quality of the outputs and also in terms of being capable of transferring colors between corresponding elements of source/target pairs, particularly small elements and details.

The results indicated that the classification based approach is a better option, not only because of its impact over the initial color assignment, but also because it enables the key modules of the method pipeline, the two-stage classification and the edge-aware relabeling. The employment of these modules culminate with the proposed method presenting results that are visually better compared to its exemplar-based counterparts.

Although not explored in this chapter, different ratios of feature weights lead to different colorization results. Since the goal of this chapter was to compare the proposed method with baseline methods in a general sense. However, in order to explore these methods to their full potential, it might necessary to tune the weights according to the image pair at hand.

Finally, the results generated by a recent learning-based algorithm (ZHANG; ISOLA; EFROS, 2016) for our input images confirmed that the colorization of the selected complex images is a difficult problem even for an algorithm of a different class. The proposed method was capable of generating comparable results to the ones by the learning-based method while allowing for some user control over the output results and showing more well defined color assignments around object boundaries/edges.

4.5.1 Limitations

The method is designed for the scope of natural images, therefore drawings and synthetic (computer generated) images in general are not expected to work properly because they lack the characteristics of natural images that are explored by the feature extraction.

The superpixel segmentation employed in the proposed method is sensitive to image scale in different ways. Images that are too small cause the applied superpixel segmentation to be highly non-uniform in shape and can compromise the maintenance of the original image structure. Also, the superpixels are not able to capture small image details leading to bleeding artifacts as mentioned in Section 4.3.1 and previously reported in (GUPTA et al., 2012).

Finally, as for the exemplar-based methods in general, the proposed method relies on the availability of source color images with similar content to the gray target.

5 CONCLUSION

In this dissertation, we presented a comprehensive review of the image colorization literature, ranging from the pioneering works to the most recent ones. A taxonomy was introduced that categorizes the reviewed algorithms based on the source of prior information used to guide their color assignments. The review included representatives from all categories and examined how the later works built upon the previous and on which aspects they aimed to improve them. The review showed that image colorization is an active research topic having many papers being published in recent years.

The review process showed that most of the exemplar-based methods in the literature have their successful result examples largely oriented to simple source/target pairs, such as landscapes, animals and simple buildings. Such images are characterized by a straightforward mapping between colors and local descriptions. Based on this observation, an exemplar-based colorization method was proposed in this work with the intent of generating plausible colorizations for a broader set of input pairs, including images that possess complexity traits such as color imbalances, colorfulness, blurred regions, multimodality, and so on.

The proposed method makes use of superpixel segmentation, which provides a faster execution as well as a greater coherence, and combines low, medium and high level features to create a superpixel local descriptor space in which metrics are more distinctive. It also features a two-stage cost-based classification which improves the accuracy of target superpixels, while also taking into account the color imbalances in the source image. Lastly, an edge-aware relabeling scheme that enforces spatial coherence while respecting the original image edge structure is also included in the proposed pipeline.

Experiments were performed to validate the design decisions from the proposed method and compare the colorization results to two baseline algorithms from the literature. The comparative analysis was performed subjecting the methods to selected simple and complex source/target pairs to visually assess and compare their colorization results. Experimental results showed that the modules that comprise the proposed pipeline have positive impact when analyzed in isolation. Moreover, result images from the proposed method showed more visual quality than the baseline methods, both within the simple images set, in which the color assignments around boundaries were improved and on the complex images, in which the proposed method transferred colors between corresponding image elements more successfully. Thus, the experimental analysis indicates that incorporating into the pipeline modules that enhance robustness to local descriptors limitations allows exemplar-based methods to handle complex image pairs with relative success and represent an improvement over previous methods. The experimental analysis also featured results from a state-of-the-art learning-based algorithm to provide a better sense of how the proposed method results compare to the modern deep learning approaches. These results indicate that exemplar-based methods can, to some extent, generate comparable results while allowing for user control over the final result.

As for future work, a few directions could be followed. The definition of a feature set is still a point of no consensus, as different methods in the literature apply different combinations of known computer vision features. A possible line of investigation may involve an experimental analysis on feature set design for image colorization considering data complexity measures in the generated feature space to determine whether there are features more suited for certain types of images. Developing techniques to perform automatic feature selection or feature weight optimization within a predefined set would also be an interesting contribution. Local descriptor synthesis techniques (such as (AL-SAHAF et al., 2015)) could also be explored in the context of colorization. Finally, to enhance our experimental analysis, it would be beneficial to perform a user study to evaluate our results, both in comparison to real images, to assess the *naturalness* of the proposed method outputs, and in comparison with the baseline results to have a less subjective mean of comparison.

REFERENCES

AL-SAHAF, H.; ZHANG, M.; JOHNSTON, M.; VERMA, B. Image descriptor: A genetic programming approach to multiclass texture classification. In: IEEE. *Evolutionary Computation (CEC), 2015 IEEE Congress on.* [S.l.], 2015. p. 2460–2467.

BAIG, M. H.; TORRESANI, L. Multiple hypothesis colorization and its application to image compression. *Computer Vision and Image Understanding*, Elsevier, v. 164, p. 111–123, 2017.

BALINSKY, A.; MOHAMMAD, N. Colorization of natural images via l 1 optimization. In: IEEE. *Applications of Computer Vision (WACV), 2009 Workshop on.* [S.l.], 2009. p. 1–6.

BUGEAU, A. *Patch-based models for image post-production.* Phd Thesis (PhD Thesis) — Université de Bordeaux, 2018.

BUGEAU, A.; TA, V.-T. Patch-based image colorization. In: IEEE. *Pattern Recognition (ICPR), 2012 21st International Conference on.* [S.l.], 2012. p. 3058–3061.

BUGEAU, A.; TA, V.-T.; PAPADAKIS, N. Variational exemplar-based image colorization. *IEEE Transactions on Image Processing*, IEEE, v. 23, n. 1, p. 298–307, 2014.

CHAMBOLLE, A.; POCK, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, Springer, v. 40, n. 1, p. 120–145, 2011.

CHARPIAT, G.; HOFMANN, M.; SCHÖLKOPF, B. Automatic image colorization via multimodal predictions. In: SPRINGER. *European conference on computer vision.* [S.l.], 2008. p. 126–139.

CHENG, Z.; YANG, Q.; SHENG, B. Deep colorization. In: *Proceedings of the IEEE International Conference on Computer Vision.* [S.l.: s.n.], 2015. p. 415–423.

CHIA, A. Y.-S.; ZHUO, S.; GUPTA, R. K.; TAI, Y.-W.; CHO, S.-Y.; TAN, P.; LIN, S. Semantic colorization with internet images. In: ACM. *ACM Transactions on Graphics (TOG).* [S.l.], 2011. v. 30, n. 6, p. 156.

COLORING Old Movies: Foes See Red, Backers See Green. 1986. Chicago Tribune <http://articles.chicagotribune.com/1986-08-29/entertainment/8603050091_1_wilson-markle-constance-bennett-laurel-and-hardy-movie>. [Online; accessed 05-April-2018].

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. ImageNet: A Large-Scale Hierarchical Image Database. In: *Computer Vision and Pattern Recognition.* [S.l.: s.n.], 2009.

DESHPANDE, A.; LU, J.; YEH, M.-C.; FORSYTH, D. A. Learning diverse image colorization. *CoRR, abs/1612.01958*, v. 1, 2016.

DESHPANDE, A.; ROCK, J.; FORSYTH, D. Learning large-scale automatic image colorization. In: *Proceedings of the IEEE International Conference on Computer Vision.* [S.l.: s.n.], 2015. p. 567–575.

FATTAL, R. Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics (TOG)*, ACM, v. 28, n. 3, p. 22, 2009.

GOFFAUX, V.; JACQUES, C.; MOURAUX, A.; OLIVA, A.; SCHYNS, P.; ROSSION, B. Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, Taylor & Francis, v. 12, n. 6, p. 878–892, 2005.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing.* [S.l.]: Upper Saddle River, NJ: Prentice Hall, 2012.

GORDON, I. E. *Theories of visual perception.* [S.l.]: Psychology Press, 2004.

GU, X.; HE, M.; GU, X. Thermal image colorization using markov decision processes. *Memetic Computing*, Springer, v. 9, n. 1, p. 15–22, 2017.

GUPTA, R. K.; CHIA, A. Y.-S.; RAJAN, D.; NG, E. S.; ZHIYONG, H. Image colorization using similar images. In: ACM. *Proceedings of the 20th ACM international conference on Multimedia.* [S.l.], 2012. p. 369–378.

HAMAM, T.; DORDEK, Y.; COHEN, D. Single-band infrared texture-based image colorization. In: IEEE. *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of.* [S.l.], 2012. p. 1–5.

HERTZMANN, A.; JACOBS, C. E.; OLIVER, N.; CURLESS, B.; SALESIN, D. H. Image analogies. In: ACM. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques.* [S.l.], 2001. p. 327–340.

HUA, M.; BIE, X.; ZHANG, M.; WANG, W. Edge-aware gradient domain optimization framework for image filtering by local propagation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* [S.l.: s.n.], 2014. p. 2838–2845.

HUANG, Y.-C.; TUNG, Y.-S.; CHEN, J.-C.; WANG, S.-W.; WU, J.-L. An adaptive edge detection based colorization algorithm and its applications. In: ACM. *Proceedings of the 13th annual ACM international conference on Multimedia.* [S.l.], 2005. p. 351–354.

HYVÄRINEN, A.; HURRI, J.; HOYER, P. O. *Natural image statistics: a probabilistic approach to early computational vision.* [S.l.]: Springer, 2009.

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 (TOG)*, ACM, v. 35, n. 4, p. 110, 2016.

IRONY, R.; COHEN-OR, D.; LISCHINSKI, D. Colorization by example. In: CITESEER. *Rendering Techniques.* [S.l.], 2005. p. 201–210.

JAIN, A. K.; FARROKHNIA, F. Unsupervised texture segmentation using gabor filters. *Pattern recognition*, Elsevier, v. 24, n. 12, p. 1167–1186, 1991.

LARSSON, G.; MAIRE, M.; SHAKHNAROVICH, G. Learning representations for automatic colorization. In: SPRINGER. *European Conference on Computer Vision.* [S.l.], 2016. p. 577–593.

LEE, S.; PARK, S.-W.; OH, P.; KANG, M. G. Colorization-based compression using optimization. *IEEE Transactions on Image Processing*, IEEE, v. 22, n. 7, p. 2627–2636, 2013.

LEVIN, A.; LISCHINSKI, D.; WEISS, Y. Colorization using optimization. In: ACM. *ACM Transactions on Graphics (ToG).* [S.l.], 2004. v. 23, n. 3, p. 689–694.

LEVINSHTEIN, A.; STERE, A.; KUTULAKOS, K. N.; FLEET, D. J.; DICKINSON, S. J.; SIDDIQI, K. Turbopixels: Fast superpixels using geometric flows. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 31, n. 12, p. 2290–2297, 2009.

LI, B.; LAI, Y.-K.; ROSIN, P. L. Example-based image colorization via automatic feature selection and fusion. *Neurocomputing*, Elsevier, v. 266, p. 687–698, 2017.

LI, B.; ZHAO, F.; SU, Z.; LIANG, X.; LAI, Y.-K.; ROSIN, P. L. Example-based image colorization using locality consistent sparse representation. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 11, p. 5188–5202, 2017.

LIU, X.; WAN, L.; QU, Y.; WONG, T.-T.; LIN, S.; LEUNG, C.-S.; HENG, P.-A. Intrinsic colorization. In: ACM. *ACM Transactions on Graphics (TOG).* [S.l.], 2008. v. 27, n. 5, p. 152.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, n. 2, p. 91–110, 2004.

LUAN, Q.; WEN, F.; COHEN-OR, D.; LIANG, L.; XU, Y.-Q.; SHUM, H.-Y. Natural image colorization. In: EUROGRAPHICS ASSOCIATION. *Proceedings of the 18th Eurographics conference on Rendering Techniques.* [S.l.], 2007. p. 309–320.

MARTINEZ-ESCOBAR, M.; FOO, J. L.; WINER, E. Colorization of ct images to improve tissue contrast for tumor segmentation. *Computers in biology and medicine*, Elsevier, v. 42, n. 12, p. 1170–1178, 2012.

MORIMOTO, Y.; TAGUCHI, Y.; NAEMURA, T. Automatic colorization of grayscale images using multiple images on the web. In: ACM. *SIGGRAPH'09: Posters.* [S.l.], 2009. p. 32.

O'DONOVAN, P.; AGARWALA, A.; HERTZMANN, A. Color compatibility from large datasets. In: ACM. *ACM Transactions on Graphics (TOG).* [S.l.], 2011. v. 30, n. 4, p. 63.

PIERRE, F.; AUJOL, J.-F.; BUGEAU, A.; PAPADAKIS, N.; TA, V.-T. Exemplar-based colorization in rgb color space. In: IEEE. *Image Processing (ICIP), 2014 IEEE International Conference on.* [S.l.], 2014. p. 625–629.

PIERRE, F.; AUJOL, J.-F.; BUGEAU, A.; PAPADAKIS, N.; TA, V.-T. Luminance-chrominance model for image colorization. *SIAM Journal on Imaging Sciences*, SIAM, v. 8, n. 1, p. 536–563, 2015.

PIERRE, F.; AUJOL, J.-F.; BUGEAU, A.; TA, V.-T. *Colociel. Solution for Image Colorization.* [S.l.], 2016.

POPURI, K. Introduction to variational methods in imaging. In: . [S.l.: s.n.], 2010.

QU, Y.; WONG, T.-T.; HENG, P.-A. Manga colorization. In: ACM. *ACM Transactions on Graphics (TOG)*. [S.l.], 2006. v. 25, n. 3, p. 1214–1220.

REINHARD, E.; ADHIKHMIN, M.; GOOCH, B.; SHIRLEY, P. Color transfer between images. *IEEE Computer graphics and applications*, IEEE, v. 21, n. 5, p. 34–41, 2001.

REINHARD, E.; CUNNINGHAM, D. W.; POULI, T. *Image statistics in visual computing*. [S.l.]: AK Peters/CRC Press, 2013.

REN, X.; MALIK, J. Learning a classification model for segmentation. In: IEEE. *null*. [S.l.], 2003. p. 10.

RUBNER, Y.; TOMASI, C.; GUIBAS, L. J. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, Springer, v. 40, n. 2, p. 99–121, 2000.

SHENG, B.; SUN, H.; CHEN, S.; LIU, X.; WU, E. Colorization using the rotation-invariant feature space. *IEEE computer graphics and applications*, IEEE, v. 31, n. 2, p. 24–35, 2011.

SIERRA, B.; LAZKANO, E.; IRIGOIEN, I.; JAUREGI, E.; MENDIALDUA, I. K nearest neighbor equality: giving equal chance to all existing classes. *Information Sciences*, Elsevier, v. 181, n. 23, p. 5158–5168, 2011.

SONG, Q.; XU, F.; JIN, Y.-Q. Radar image colorization: Converting single-polarization to fully polarimetric using deep neural networks. *IEEE Access*, IEEE, v. 6, p. 1647–1661, 2018.

TAI, Y.-W.; JIA, J.; TANG, C.-K. Local color transfer via probabilistic segmentation by expectation-maximization. In: IEEE. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.], 2005. v. 1, p. 747–754.

WANG, X.-H.; JIA, J.; LIAO, H.-Y.; CAI, L.-H. Affective image colorization. *Journal of Computer Science and Technology*, Springer, v. 27, n. 6, p. 1119–1128, 2012.

WELSH, T.; ASHIKHMIN, M.; MUELLER, K. Transferring color to greyscale images. In: ACM. *ACM Transactions on Graphics (TOG)*. [S.l.], 2002. v. 21, n. 3, p. 277–280.

YANG, C.; ZHANG, L.; LU, H.; RUAN, X.; YANG, M.-H. Saliency detection via graph-based manifold ranking. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2013. p. 3166–3173.

YATZIV, L.; SAPIRO, G. Fast image and video colorization using chrominance blending. *IEEE transactions on image processing*, IEEE, v. 15, n. 5, p. 1120–1129, 2006.

ZHANG, R.; ISOLA, P.; EFROS, A. A. Colorful image colorization. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 649–666.

ZHANG, R.; ZHU, J.-Y.; ISOLA, P.; GENG, X.; LIN, A. S.; YU, T.; EFROS, A. A. Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999*, 2017.

ZHENG, Y.; ESSOCK, E. A. A local-coloring method for night-vision colorization utilizing image analysis and fusion. *Information Fusion*, Elsevier, v. 9, n. 2, p. 186–199, 2008.