IRSTI 50.05.13

A.M. Ashiraliyev[1],
Guo Ling[2] (orcid 0000-0001-8227-3528) – main author

[1]Master, [2]PhD, Associate professor
Nanjing University of Science and Technology, Nanjing, China
E-mail: [1]aiadil12@mail.ru

THE SCORE IMAGE PROCESSING AND CORRECTION

Abstract. With the development of electronic technologies, scanned or photographed paper documents are the main data for processing by computer applications; musical scores are taken as research in the work. The article presents methods of data processing by bilateral filtering and a noise reduction algorithm in order to improve the quality of images of musical scores and/or technical documentations.

Keywords: musical score, image processing, linear and non-linear filtering, binarization process, image recognition.

Introduction. The object of the research is images of musical scores obtained by using mobile terminals such as mobile phones or tablets. Image quality has an important impact on the complexity of algorithm design, as well as on the efficiency and accuracy of recognition. Due to the quality of the paper itself, score images often have the following problems: The color tone is yellowish and noise is present, which is caused by differences in paper quality, poor user data storage, and intrinsic noise of the electronic equipment used for taking photo, as shown in Figure 1.

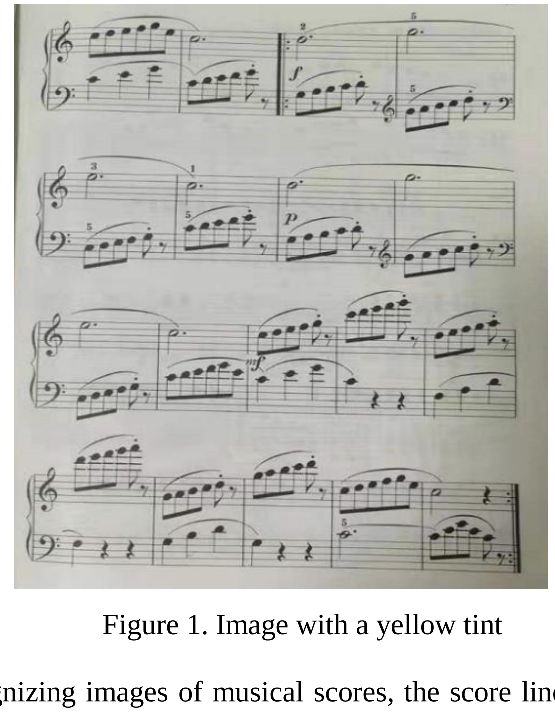

Figure 1. Image with a yellow tint

For recognizing images of musical scores, the score line is one of the most important musical symbols and a benchmark for recognizing the pitch of notes. In actual acquired images, both tilt and shear deformation are often present, and in this case only the slope has been corrected. Thus, the purpose of image processing is to improve image quality by removing noise, as well as to perform automatic threshold binarization of the image.

To eliminate problems with noise, color cast and spectral line tilt in the score image, it was necessary to first perform pre-processing.

When the score is actually captured, it is affected to a greater or lesser extent by distortions inherent in the imaging device and the environment, so that the resulting digital image contains noise and appears distorted. In view of this situation, it was necessary to perform noise reduction processing on the score image in order to restore the authenticity of the image, also it was necessary to reduce the effect of noise on subsequent operations, as well as ensuring the efficiency and reliability of post-processing.

The existing methods for smoothing and filtering images mainly include two aspects: linear filtering and non-linear filtering [1]. Linear filtering includes: Gaussian filtering, mean filtering. Nonlinear filtering includes: bilateral filtering, median filtering.

The core of Gaussian filtering is the use of a two-dimensional Gaussian function. After the Gaussian function passes the Fourier transform, the performance of the Gaussian function in the frequency domain is still its function [2]. Therefore, the Gaussian function can generate a low pass filter with smoothing performance in the frequency domain. Smoothing filtering can be achieved by multiplying the product in the frequency domain, and also the smoothing effect is different for different deviations. Mean filtering used to locally average the signal, and the average value is used to represent the gray value of the pixel, as shown in Figure 2(a). For general mean filtering of averages, as shown in fig. 3(b), the calculation of the average value will lead to the “blurring” of information about the edges and about objects in the image, thereby losing a lot of data. Gaussian filtering cannot avoid this situation, as shown in Figure 3(c).

$$\begin{pmatrix} 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \\ 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \\ 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \\ 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \\ 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \end{pmatrix} \begin{pmatrix} 6.96e-08 & 2.81e-05 & 2.08e-04 & 2.81e-05 & 6.96e-08 \\ 2.81e-05 & 0.01133 & 0.08373 & 0.01133 & 2.81e-05 \\ 2.08e-04 & 0.08373 & 0.6187 & 0.08373 & 2.08e-04 \\ 2.81e-05 & 0.01133 & 0.08373 & 0.01133 & 2.81e-05 \\ 6.96e-08 & 2.81e-05 & 2.08e-04 & 2.81e-05 & 6.96e-08 \end{pmatrix}$$

(a) average smoothing matrix (b) Gaussian smoothing matrix

Figure 2. Smoothing matrix

(a) original image (b) average smoothing result (c) Gaussian smoothing result

Figure 3. Diagram of the smoothing effect

Under normal circumstances, the image of the score is mainly divided into notes and background, and the colors of both elements are very different, as shown in Fig. 4. If a linear filter is used, then the transition between them will be smoothed, and the result of such processing will have a certain effect on the subsequent binarization and spectral line positioning. Both bilateral filtering and median filtering are non-linear filters that preserve the edges of the image. Median filtering has a good suppression effect on salt-and-pepper noise, on the other hand, median filtering will destroy musical notes and spectral lines, resulting in a lot of information being lost in musical scores, as shown in Figure 5. Bilateral filtering uses a weighted average method, which uses a weighted average of the luminance values of the surrounding pixels to represent the intensity of a certain pixel, and the weighted average based on the Gaussian distribution. Most importantly, the weight of bilateral filtering not only considers the Euclidean distance of the pixels, but also considers the radiation differences in the pixel range. The weight calculation formula for bilateral filtering is as follows

$$w(i,j,k,l)=\exp(-\frac{(x-k)^2+(y-l)^2}{2\sigma_d^2}-\frac{(I(x,y)-I(k,l)^2)}{2\sigma_r^2}) \tag{1}$$

Where: σ_d, σ_r are smoothing parameters; (x, y) — smoothing pixel coordinates, (k, l) — neighboring pixel coordinates; I (x, y),I (k, l) — grayscale values corresponding to pixel coordinates (x, y) and (k, l); w(i, j, k, l) is the calculated weight. The bilateral filter smoothing formula is as follows：

$$I_D(x,y)=\frac{\sum_{k,l} I(k,l)w(x,y,k,l)}{\sum_{k,l} w(x,y,k,l)} \tag{2}$$

Where: w(x, y, k, l) is calculated by formula 1, the range of values kl depends on the size of the neighborhood, I (k, l) is the pixel coordinate. The gray value corresponds to the scale (x, y). In view of the situation described above, bilateral filtering is used to remove noise from the musical score. By setting σ_d = 0.75, σ_r = 0.5, neighborhood size 25 * 25 in formula 1, the following results are obtained in Figure 6. It can be seen from the figure that due to the image feature, the bilateral filter has incomparable advantages compared to the average filtering, which can preserve edges, reduce noise, and smooth without removing notes and spectral lines.

Figure 4. Original image

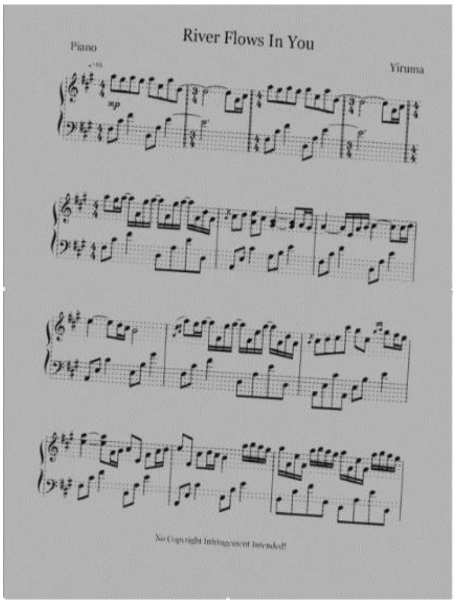

Figure 5. Median filtering results

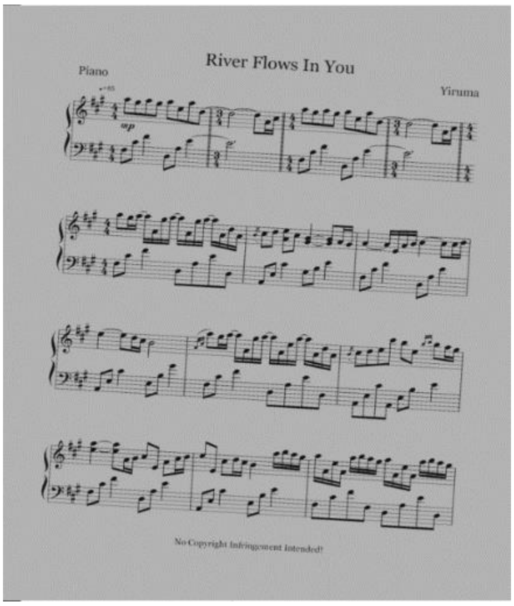

Figure 6. Bilateral filtering results

The information in the score image is mainly focused on notes and lines, while the background does not provide any information. On the other hand, when shooting music nodes, the background of the resulting image may not be white due to the ambient light or color of paper, and if the overall brightness is low, the contrast of the foreground and background in the image may not be clear. Considering the above two aspects, an operation was performed to improve the image of the musical score by increasing the contrast between the background and foreground, and avoided the unnecessary noise introduced by directly binarizing the image. Image enhancement is one of the widely used image processing techniques. The purpose of image enhancement is to make the original blurry image sharp or highlight some important features with image enhancement technology. The existing image enhancement methods are mainly divided into two categories: the frequency domain method and the spatial domain method. The frequency domain method includes: filtering in the frequency domain, homomorphic filtering, and the spatial domain method includes: grayscale transformation, histogram correction [2]. Frequency-domain algorithms such as homomorphic filtering are inefficient at enhancing detail; spatial domain-based algorithms such as grayscale conversion are difficult to use because it is difficult to achieve a good trade-off between detail enhancement and noise reduction. The situation where the overall brightness of the image is low due to insufficient lighting while taking photo of the score, a histogram equalization method was used to solve this problem. This method is effective for images with too bright or too dark backgrounds and foregrounds. Histogram equalization basically transforms the statistical information of an image, in order to achieve the contrast enhancement effect, the following formula was used:

$$p(L)=\frac{n_L}{width*height},L=0,1,2.....,M \tag{3}$$

Where: height and width are the height and width of the image, and n_L is the number of pixels that have a gray value equal to L in the image, that is, the frequency, p(L) is the frequency of occurrence of L, and M is the maximum gray level. After receiving the statistical information, the conversion formula looks like this:

$$S_k=T(L)=(M-1)\sum_{j=0}^{L}p(j)=\frac{M-1}{width*height}\sum_{j}^{L}n_j,L=0,1,2,...M \tag{4}$$

Where: p(j), n_j are calculated by formula 3, S_k is the output value of the display with the gray level L in the original image. The image histogram equalization is calculated using formulas 3 and 4, the result is shown in Fig. 7.

Image recognition usually requires the use of an image segmentation algorithm to distinguish the object of interest from other objects. On this basis, the pixels of the foreground object (binarization) have been checked and colored, and the background object is white. The goal of binarization is to filter out useful data, thereby reducing the amount of computation required to process the data.

The objects of research are mainly musical symbols and spectral lines, the rest of the information has a secondary nature and, accordingly, they were deleted.

For a grayscale image f (x, y), the binarization process is to find a suitable threshold value T and convert it into a binarized image g(x, y) as shown in the following formula：

$$g(x,y)=\begin{cases}255 & f(x,y)>T\\ 0 & f(x,y)\le T\end{cases} \tag{5}$$

Figure 7. The result of the enhanced image

According to the principle used to calculate the threshold T, the most commonly used binarization methods can be divided into the global threshold method [3], the local threshold method [4], the maximum inter-category variance method [5] and the local text information method [6]. Due to differences in paper quality and light differences while taking a photo, it is difficult to achieve a good balance between automation, computational complexity, and effects with the global threshold method, local threshold method, and local text information method. Given that there is no complex foreground and background in the image of the score, the foreground mainly consists of notes containing a lot of musical information, while the background often has a transitional smooth color, and they have strong contrast after image enhancement, as shown in the figure. 7. In accordance with the above characteristics of the score image, the maximum inter-category variance method was used to binarize the score image[6]. The maximum inter-category variance method is an adaptive thresholding method proposed by the Japanese scientist Otsu in 1979[5], which can automatically calculate the segmentation threshold according to the characteristics of the gray distribution in the image, and separate the image into background and foreground. The algorithm includes the following steps:

1. Counting the number of pixels with different gray levels in an image, that is, the frequency of each gray level n_i, $0 \le i \le M$. M is the maximum gray level, which typically equals 255.
2. Computing the frequency p_i, $0 \le i \le M$ of each grayscale. Calculated as follows:

$$p_i = \frac{n_i}{width * height}, i = 0,1,2.....,M \tag{6}$$

Where: width and height are the height and width of the image, n_i is the number of pixels whose gray value is equal to i in the image, i.e. the frequency, p_i is the frequency of occurrence of i, and M is the maximum value of the gray level.

3. Setting the classification threshold δ , 0 ≤ δ ≤ M. The grayscales of a pixel located in the range [0, δ] are the background pixel, and the grayscales of the pixel located in the range [δ, M] are the foreground pixels. Iteration from δ = 0 to δ = M calculates the inter-category variance at various δ, the formula is as follows:

$$v_2 = w_0 * v_0 + w_1 * v_1 \quad (7)$$

$$\sigma = w_0 * (v_0 - v_2)^2 + w_1 * (v_1 - v_2)^2 \quad (8)$$

In this formula, w_0 and v_0 are the proportion and average of the background gray pixels in the entire image, w_1 and v_1 are the proportion and average of the foreground gray pixels in the entire image, respectively, and v_2 is the proportion of the overall average of the image. Shades of gray, σ - variance between classes (foreground and background).

4. Comparing all σ , δ, the maximum threshold value was obtained as the global image threshold, and then using formula 5 to perform the binarization operation, the following result was obtained in Fig. 8.

Figure 8. Graph of binarization results

Conclusion. The research work mainly dealt with the problems of noise, color tone in the image of the score, and also performed image noise reduction, enhancement and binarization. A bilateral filtering algorithm was used to denoise the image of the score, which effectively preserves information about the borders of the image. To solve the problem of low overall brightness of the captured image, we used histogram equalization, which effectively improves the contrast between background and foreground, thereby facilitating image binarization based on maximum interclass variance. The presented research methods allow to restore and/or improve the quality of images of musical scores or technical documentations.

References

1. Blanes AR, Bisquerra AF. Camera-based Optical Music Recognition using a Convolutional Neural Network [C]//2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2017, 2: 27-28.
2. Mo Deju, Liang Guanghua. Digital Image Processing [M]. Beijing University of Posts and Telecommunications Press, 2010.
3 Ng K, Cooper D, Stefani E, et al. Embracing the Composer: Optical Recognition of Handwritten Manuscripts[C]//ICMC. 1999.
4. Niblack W. An introduction to digital image processing, 1986[C]//G. Leedham, C. Yan, K. Takru, J. Tan and L. Mian. Comparison of Some Thresholding Algorithms for Text/Background Segmentation in Difficult Document Images. In Proceedings of the Seventh International Conference on Document Analysis and Recognition. 2003: 859-864.
5. Otsu N. A threshold selection method from gray-level histograms[J]. IEEE transactions on systems, man, and cybernetics, 1979, 9(1): 62-66.
6. Fornés A, Lladós J, Sánchez G. Primitive segmentation in old handwritten music scores[C]//International Workshop on Graphics Recognition. Springer, Berlin, Heidelberg, 2005: 279-290.

Material received 17.02.22.

А.М. Әшіралиев, Гуо Лин

Нанкин ғылым және технология университеті, Нанкин қ., Қытай

НОТАЛЫҚ КЕСКІНДІ ӨҢДЕУ ЖӘНЕ ТҮЗЕТУ

Аңдатпа. Электрондық технологиялардың дамуымен сканерленген немесе фотосуретке түсірілген қағаз құжаттар компьютерлік қолданбалармен өңдеу үшін негізгі деректер болып табылады, жұмыста зерттеу ретінде музыкалық партитуралар алынды. Мақаладамузыкалық партитуралардың және/немесе техникалық құжаттаманың кескіндерінің сапасын жақсарту үшін екі жақты сүзгілеу және шуды азайту алгоритмі арқылы деректерді өңдеу әдістері ұсынылған.

Тірек сөздер: музыкалық партитура, кескінді өңдеу, сызықтық және сызықты емес сүзгілеу, бинаризация процесі, кескінді тану.

А.М. Аширалиев, Гуо Линь

Нанкинский университет науки и технологий, г. Нанкин, Китай

ОБРАБОТКА И КОРРЕКЦИЯ ИЗОБРАЖЕНИЯ НОТНЫХ ПАРТИТУР

Аннотация. С развитием электронных технологий, сканированные или сфотографированные документации в бумажных носителях являются основными данными для обработки компьютерными приложениями, в работе в качестве исследования приняты нотные партитуры. В статье приводятся методы обработки данных двусторонней фильтрацией и алгоритмом шумоподавления с целью улучшения качества изображений нотных партитур и/или технических документаций.

Ключевые слова: музыкальная партитура, обработка изображений, линейная и нелинейная фильтрация, процесс бинаризации, распознавание изображений.