International Journal of Engineering

Super-resolution of License-plates Using Weighted Interpolation of Neighboring Pixels from Video Frames

K. Mehrgan*, A. R. Ahmadyfard, H. Khosravi

Faculty of Electrical Engineering and Robotics, Shahrood University of Technology, Shahrood, Iran

P A P E R I N F O	*A B S T R A C T*

Recognizing the license plate from a set of low-resolution video frames using an Optical Character Recognition system (OCR) is a very challenging task. OCR systems fail to properly work in this condition. The use of high-quality cameras is a costly solution to this situation. To overcome this problem, we propose a weighted interpolation method that enhances the resolution of the license plate, using consecutive frames of a video. For this purpose, first, we register the low-resolution video frames of the license plate to the reference license plate in two steps. In the first step, a coarse registration is performed by matching the SURF features. Then a fine registration on the license plate region is performed using the phase correlation technique. After registration, the reference image of the license plate is up-sampled to the desired scale. We propose a method for estimating the intensity of pixels in the up-sampled image with an unknown value. In this method, we use a weighted averaging strategy to estimate the intensity of unknown pixels using the neighboring pixels from video frames. The obtained super-resolution is suitable for OCR. Experimental results show that applying the proposed method on low-resolution frames of the license plate, improves the quality of the license plate significantly.

doi: 10.5829/ije.2020.33.05b.33

1. INTRODUCTION

Surveillance cameras are becoming an essential and widespread device in today's life. One of the applications of surveillance cameras in traffic control systems is the automatic identification of license plates by OCR systems. OCR systems require an image of proper quality to recognize the license plate. Providing a high-quality image is sometimes costly and is not always possible. There are different reasons to have a low-quality image of the license plate, the camera CCD may have a low-resolution or the camera is set to a low-resolution to save storage media and bandwidth. Sometimes an adverse atmospheric condition or considerable distance between vehicle from the camera prevents recording a high-resolution frame. Vehicle identification becomes even more important when trying to identify it at the crime scene. To accurately identify the vehicle in a traffic control system, the

resolution of the license plate must increase. In this paper, we propose a new method to increase the resolution of a license plate given a sequence of low-resolution frames. The advantage of the proposed method is its simplicity and proper performance for constructing the super-resolution image. The main question of this research is how to make a high-resolution image from a set of low-resolution images taken from the license plate of a vehicle. Our goal is to make the output image more comprehensible to humans. Moreover, we expect an OCR system can recognize the characters in the license plate accurately. Due to the poor quality of the input images, any image processing operations on them are associated with an error. One of the challenges we face is to choose a method that is less error-prone in dealing with such low-quality images.

In Section 2, we briefly review related works. In Section 3, we explain the basic idea with a simple example and describe in detail the steps of license plate

*Corresponding Author Email:: *komail.mehrgan1994@gmail.com*
(K. Mehrgan)

super-resolution. In Section 4 the results of experiments are reported. The paper is concluded in Section 5.

2. RELATED WORKS

Super-resolution methods are divided into two general groups. 1) Single-frame super-resolution 2) Multi-frame super-resolution. As multi-frame methods use more information for reconstructing a super-resolution image in comparison with single-frame methods, the performance of multi-frame methods is superior. However, in many cases, only a single low-resolution image is provided from the scene. A single-frame method tries to add details to the output image using either content of the input image or a set of multi-resolution images given in an external database. Image interpolation is one of the most basic approaches for this purpose.

In [1] a method was proposed for super-resolution in which common patterns are extracted from an external database using neural networks. The multi-frame super-resolution has been studied since 1984. Tsai [2] introduced a method for multi-frame super-resolution in the frequency domain. Frequency-based super-resolution methods fail in realistic situations where the geometric transformation between frames is not as simple as transmission. For this reason, the researchers have paid attention to develop the methods in the spatial domain. One of the early methods introduced in the spatial-domain was based on interpolation [3]. Iterative Back Projection (IBP) algorithms [4] are among the early methods for spatial-based SR. These methods attempt to estimate the value of each pixel in the output image using the value of corresponding pixels from different LR frames that have small displacement or different blurring conditions. Another group of iterative methods is based on the concept of Projection onto Convex Sets (POCS) [5] These methods require accurate motion estimation of a subpixel in video frames to interpolat the unknown pixels. Recently, the concept of probabilistic registration   has been introduced [6]. These approaches perform well without the need for accurate motion estimation. Also, these are a number of multi-frame super-resolution methods based on the Maximum a Posteriori (MAP). A critical issue of the MAP-based algorithms is the choice of the prior model for the desired solution [7]. Learning-based methods have recently been introduced to solve the problem of super-resolution. In these methods, the image details are improved by recovering the high-frequency information from the training samples. Hertzman et al. [8] introduced a method for obtaining high-frequency image information using a training database. Their method has two steps. 1) Training, and 2) super-resolution image reconstruction. There are a few works that address the problem of super-resolution for the license plate. The aim is to create a super-resolution version of a vehicle license plate using a single frame or several frames from the license plate in the scene. In [9], a suitable method is presented to solve this problem. It consists of six steps: initializing, tacking, registration, reconstructing, post-processing and recognition. Briefly, they align the car's license plates in video frames using optical flow techniques and then increase the resolution of the image using the nearest neighbor pixel information. The major application of the super-resolution in license plate images is to improve character recognition. Some approaches address this problem [10, 11]. In [12] a method based on the combination of super-resolution and a deep neural network has been proposed to identify license plate from the low-resolution video. In [13] the concept of semantic super-resolution was introduced. Using this concept, a method for recognizing the characters of the license plate from extremely low-resolution videos was proposed.

3. PROPOSED METHOD

Given a sequence of frames including a vehicle, the aim is to reconstruct a super-resolution version of the license plate image. We select a reference frame from the video, then its neighboring frames are coarsely registered to the reference. Then the license plate is detected in the reference frame, so the plate region is extracted from all frames. A fine registration of the neighboring plates on the reference plate is performed in the next step. The reference plate is then up-sampled to the desired size (usually up to three times of the original size). The intensity of pixels with an unknown value in the super-resolution plate is estimated using a new strategy using the neighboring pixels from the neighboring frames as will be explained later.

Figure 1 shows the up-sampling of a small image patch 2×2 by scaling factor 2. The values of the new pixels are estimated using the neighboring pixels in the video frames. The use of information from the neighboring pixels to a new pixel in all video frames in super-resolution is the main difference between the proposed method and the ordinary interpolation
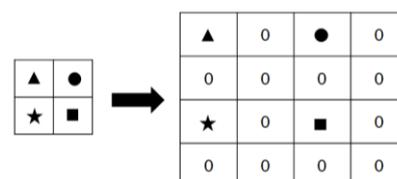

Figure 1. Extend the image matrix by adding zero rows and columns

approach. It is important since a sub-pixel shift of the plate regions causes the value of corresponding pixels changes. Hence using pixel values from the license plate in video frames provides more information in comparison to using a single frame. Different information could be integrated into a single image.

Using interpolation, the unknown pixel values are estimated when the surrounding are known. As can be seen in Figure 2, ordinary interpolations are always associated with the loss of high-frequency information while it is not the case for the proposed interpolation.

The challenge is to keep high-frequency information during the interpolation process. Indeed, the high-frequency information may exist on the scene, but it is not recorded in a frame as the number of camera sensors is limited (Figure 3b). Therefore, consecutive frames may contain different information which we aim to use all this information to build a single high-resolution image.

3. 1. Illustration
In Figure 3, we provide a simple example to illustrate the proposed interpolation in comparison with the ordinary interpolation. As shown in Figure 3-b, due to the limited number of sensors in camera, each of the captured frames includes some information from the scene. In this example, it was assumed that an object is imaged using three horizontal sensors and moves horizontally. Three frames of the image are depicted (Figure 3). The middle frame has moved a half-pixel to the left and the bottom frame has moved a half-pixel to the right, respect to the top frame (Figure 3a). Figure 3-d shows that using frames around the reference frame can produce a better and more realistic image of the object while none of the frames alone could properly describe the object. Also, interpolation on a single-frame blurs the image since it averages the value of neighboring pixels within the image (Figure 3-c).

A challenge with the proposed interpolation is that the displacement of an object (license plate) in successive frames must be estimated accurately. In many registration methods, the error for the estimated displacement is at least one pixel while we need sub-pixel displacement estimation for the proposed interpolation.
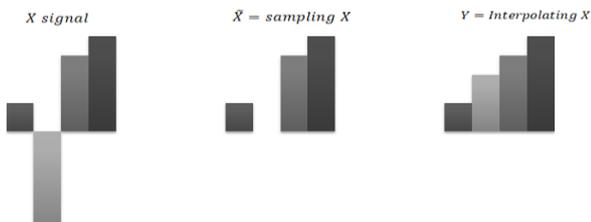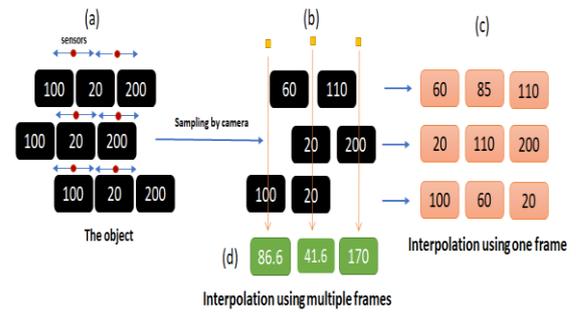

Figure 3. A simple example to illustrate single-frame interpolation vs. multi-frame interpolation. a) The object which it was imaged, b) Sampling by camera sensors, c) interpolating using one frame. The middle pixel is the average of the neighboring pixel, d) multi-frame interpolation. Green pixels are the weighted average of the pixels that hit the yellow line.

3. 2. Super-resolution of License Plate
The super-resolution of the license plate consists of two main parts: estimating the translation, rotation, and magnification of the images relative to the reference image (registration step); and super-resolution reconstruction using the multi-frame interpolation.

3. 2. 1. Registration Steps
Many methods have been proposed for image registration. These methods are generally divided into two main categories: feature-based and intensity-based. The registration step is very crucial for the performance of the super-resolution method. We need to perform the frame registration in two steps. In the first step, two frames of the vehicle are registered coarsely. After this stage, the plate region is extracted. In the next step, the license plate regions are registered to the reference frame with sub-pixel accuracy which is required for super-resolution. In this article, the license plates are cropped manually. But we're trying to find a way to crop them automatically. This is a challenge for us because of the low resolution of the images.

a) Coarse Registration: We use the feature-based registration approach to align the images to the reference image, which usually is a frame in the middle of the video. For this, we extract SURF features from the reference frame and other frames, and align all frames to the reference frame using these features. It should be noted that we align the whole images rather than the license plate area as this can improve the stability of alignment as more features are involved. This has two advantages. First once all images have been aligned to the reference image, we need to detect the license plate area only on the reference frame because the license plate area is registered to the reference frame. Second, if the registration is performed using the license plate region, the algorithm may not find enough feature points due to the small size of the

Figure 2. The ordinary interpolation fails to estimate omitted samples

plate on the frames. Figure 4 shows the coarse registration using the SURF features which aligns two frames of video.

b) Fine registration: After coarse registration, the license plate area is detected on the reference frame and the same area is extracted from the other aligned frames. As mentioned earlier we need to precisely align all the license plates on the reference frame. The registration step must be performed once again, this time for the cropped plate region. For fine registration, the transformation matrix which maps each plate region on the reference plate is estimated using:

$$T = \begin{pmatrix} s \cdot cos\theta & -s \cdot sin\theta & 0 \\ s \cdot sin\theta & s \cdot cos\theta & 0 \\ \Delta_x & \Delta_y & 1 \end{pmatrix} \quad (1)$$

where S is the scaling factor, θ is rotation angle and Δ_x, Δ_y denote the translation along x, y dimensions respectively. The phase-correlation [14] is one of the well-known approaches for this task. The phase correlation is a method that is implemented in the frequency domain using the fast Fourier transform. This method aims to maximize the correlation between two given frames by estimating a transformation that maps one frame to another.

3. 2. 2. Multi-frame Interpolation   In this section the procedure for reconstruction of super-resolution is explained (Figure 5). First, we up-sample the reference plate to the desired scale. The pixels of the other frames are mapped onto the up-sampled image according to Δ_x and Δ_y from Equation (1). Now with the up-sampled image and the pixels that have been mapped on it (Figure 1), we try to estimate the unknown pixels of the up-sampled image. The simplest way for estimating the unknown pixel value is to set it to the average intensity of the selected neighbors. This idea is similar to the idea proposed in [15].
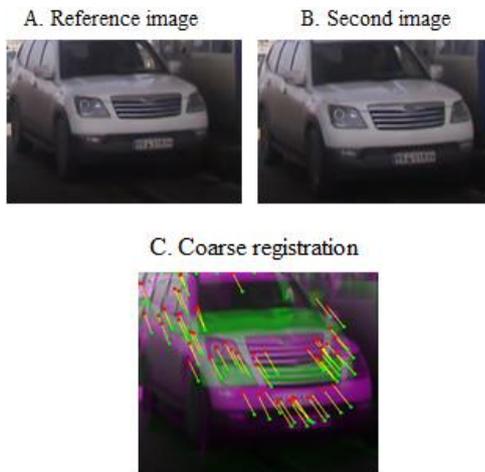
In this paper, we suggest using a weighted averaging using a number of the closet neighbors (3 to 5 neighbors) to estimate the unknown pixel. The weight for the influence of each neighbor is determined using Equation (2).

$$w_n = e^{\frac{-Distance_n}{\mu}} \quad (2)$$

$$\tilde{w}_n = \frac{w_n}{\sum_{i=1}^k w_i} \quad (3)$$

$$p = \sum_{n=1}^k \tilde{w}_n r_n \quad (4)$$

Then using Equation (4) the unknown pixel value is estimated. As we can see, the weight specialized to each neighbor is opposite to its distance and the weights are normalized using Equation (3). $Distance_n$ in Equation (2) is the Euclidean distance between the n-th neighbor from the pixel for which the intensity estimation is performed. r_n in Equation (4) is the itensity for the $n-$th neighbor to the considered pixel.

3. 2. 3. Practical Considerations   In ideal conditions, such as a negligible error in the registration step, the non-motion-blurring of some frames the proposed algorithm applies to all frames obtained from the license plate. Based on numerous experiments and considerations, we observed that in order to improve the quality of the super-resolution image as well as the speed of the algorithm, the following considerations are suggested in the proposed method. Also, a flowchart of the proposed method is shown in Figure 6.

a) The involved frames in super-resolution: It was explained in Section 3.1 that the use of consecutive frames can help us build a better picture of reality (license plate). Theoretically, using more frames would increase the quality of the super-resolution, but in practice, it is not true. We conducted an experiment to illustrate the similarity between the neighboring frames to the reference frame. We used SSIM as the similarity measure. This measure is decreased as we move away from the reference frame (Figure 7).

The considerable changes in the illumination condition and the geometric transformation between the reference frame and its neighbor justify this fact.
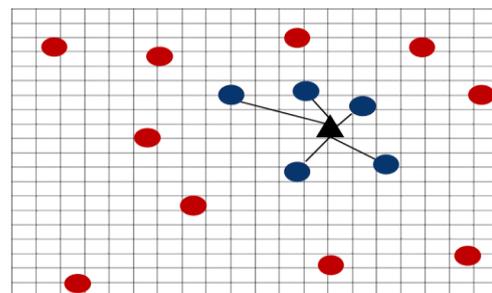

Figure 4. Coarse registration for two frames
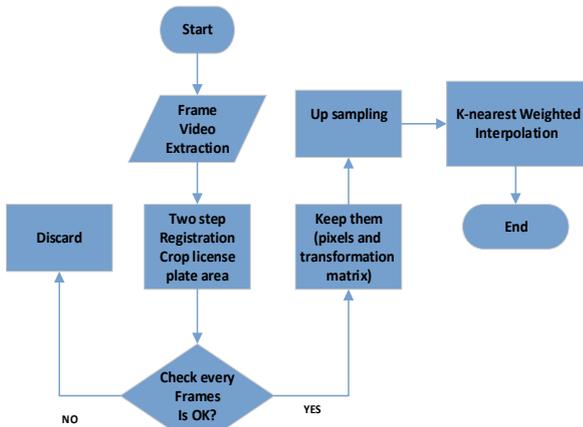

Figure 5. Multi-frame interpolation

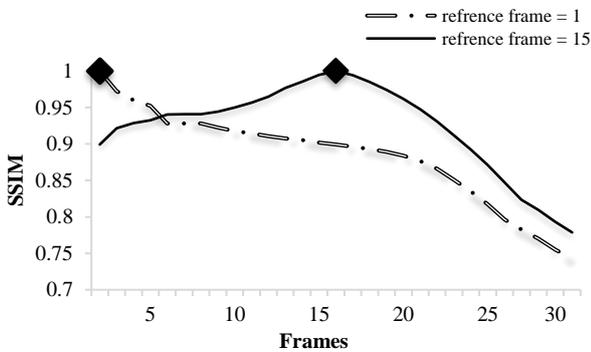**Figure 6.** Flowchart of the proposed method

Figure 7. The similarity of the images to the reference image

Experiments show that using 5 to 7 frames provides the most appropriate result. The second practical consideration is to choose the proper frame as the reference from the video frames. The SSIM decreases as the distance from the reference image is reduced. So, it makes sense that if we choose the reference image as one of the middle frames in the video stream, the super-resolution performance improves. As a result of vehicle or camera motion, the license plate in some frames is blurred. Using this frame in super-resolution causes a problem. Removing these frames from the list of neighboring frames to the reference frame improves the super-resolved result.

As for the blurred frames, high-frequency components are lost, by measuring the high-frequency components for the license plate in the neighboring frames, we detect the blurred frames. The energy for the high-frequency components is measured using the 2D Fourier transform, this value is divided by the total energy of the plate to normalize it. Among the neighboring frames, those with low normalized high-frequency energy are excluded from the list.

The registration of some frames is subject to intolerable error. It is possible due to the low-resolution

of the frames. We exclude the frames with registration errors from the list of contributing frames. For this purpose, the movement of the frames related to the reference frame along the rows (Δ_x) and columns (Δ_y) are analyzed. If these values ($\Delta x, \Delta y$) for a frame significantly differ from those in the previous and the next frames, it is subject to miss-registration. These frames are not employed for the super-resolution.

In Equation (4) when weighted interpolation is performed, we also exclude those pixels whose values are considered as the outliers, i.e. very different from the neighboring pixels.

b) Reducing the time complexity: To estimate the pixel values in the final image, it is necessary to find the k-nearest neighbor pixels from all pixels of whole frames. Finding the distance from the under reconstruction pixel to all pixels from all frames is a very time-consuming task. To speed up this task, we reduce the search space and search for nearest neighbor pixels in all frames only within the square defined by eight neighbor pixels in the reference image (see Figure 8). This reduces a significant number of operations and makes our algorithm faster. For example, suppose we want to increase an image of size 20×40 with 10 consecutive frames. The number of operations needed to find the distances from one pixel of the final image without considering neighbors is 8000 operations ($20 \times 40 \times 10$) and if we only search in the neighbors, it is 90 operations (9×10).

4. EXPERIMENTAL RESULTS

We designed two experiments to evaluate the performance of the proposed method. We collected several videos taken from a vehicle in which either vehicle or camera moves.

a) Evaluation using SSIM criterion: In the first experiment, we down-sampled the frames of each video and used the proposed super-resolution method to re-scale it to the original size (Figure 9). The output plate image for each video is compared to the plate in the reference frame of the original video. We used SSIM as the criteria to evaluate the proposed method.
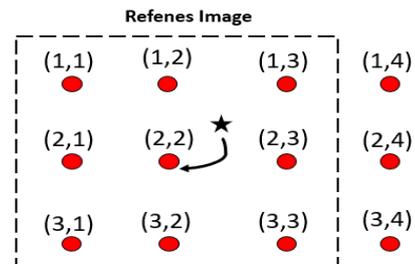

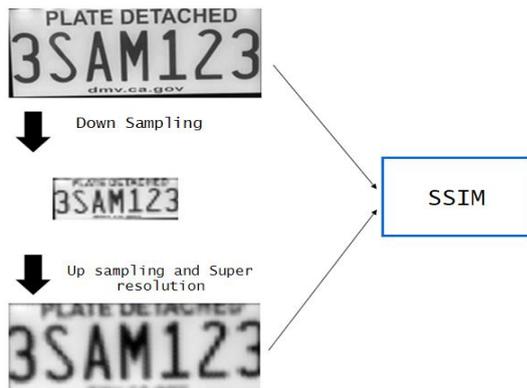**Figure 8.** Demonstration for reducing the time complexity using the proposed approach

Figure 9. Measuring the SSIM criterion using the produced super-resolution and the reference image

Figure 10 shows this criterion for the proposed muli-frame and ordinary single-frame interpolation applied on eight videos. As seen the proposed method outperforms the ordinary interpolation method.

The SSIM criterion can also be used as a quantitative criterion for the evaluation of the method. To determine the range of the scaling factor for proper super-resolution in the proposed method, we set an experiment. In this experiment for different scaling factors, we applied the proposed method on a set of license plates and for each case, the SSIM criterion was determined. Figure 11 shows for the proposed method the scaling factor can increase up to three or four while for the scaling factors larger than four the similarity
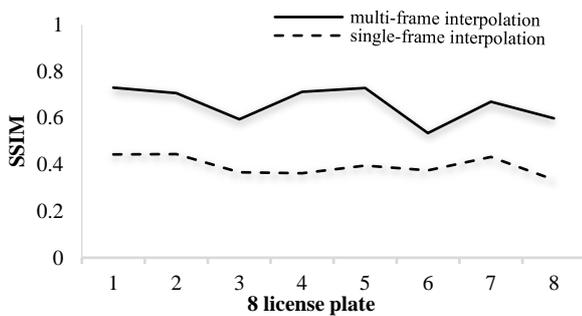
criterion decreases considerably. The result of this experiment has been reported for an average of videos.

To investigate how the number of neighboring pixels involved in the estimation affects the quality of the super-resolution image, we conducted an experiment with a single video of license plate. Figure 12 shows how the number of involved neighbors affects the similarity criterion. As can be seen, 5 to 9 neighbors provide the best result.

In Figure 13, we compare the performance of proposed weighted averaging against normal averaging for different number of involved neighbors. In this experiment, the SSIM has been determined for the super-resolution of ten license plates. The result in this experiment has been averaged on eight videos.

To study the role of parameter μ in the distance measure (Equation (2)), we conducted an experiment using eight license plate videos. Figure 14 reports the SSIM criterion against the μ parameter.

b) Plate recognition using OCR: In the second experiment, we show how the proposed super-resolution improves the performance of license plate recognition. We used Sapta [16] plate recognition software (a Persian license plate) in this study. As illustrated in Figure 15, none of the original frames were recognized by the software, after applying the super-resolution some characters were recognized correctly. In Table 1, we reported the recognition rate for the plate characters before and after super-resolution.

Figure 12. The effect of the number of geometric neighbors on SSIM

Figure 10. SSIM criterion for super-resolution using the proposed multi-frame and ordinary single-frame interpolation
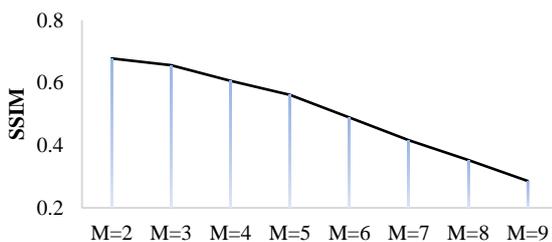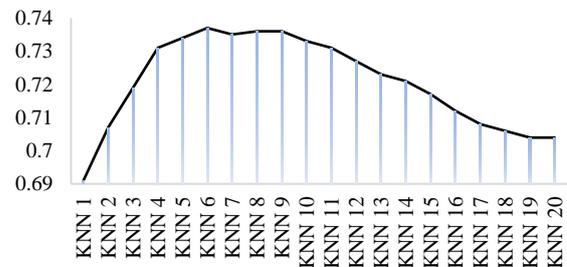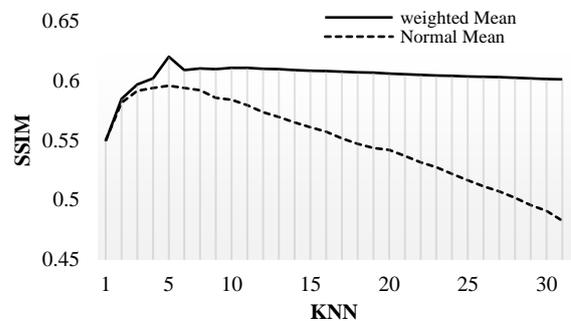

Figure 13. SSIM criterion for super-resolution of ten license plates using the proposed weighted averaging against normal averaging for different number of involved neighbors
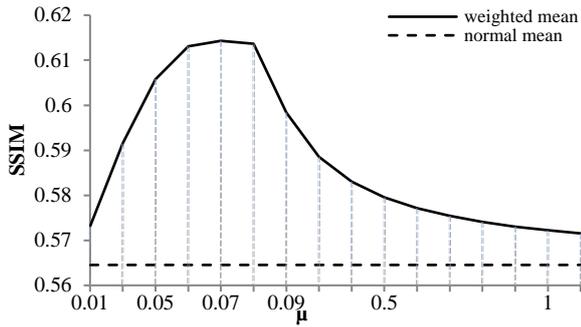

Figure 11. The SSIM criterion for different scaling factors

Figure 14. The effect of µ on distance Equation (2) on the SSIM

Figure 15. The license plate recognition using OCR

Figure 16. Some examples from the license plate images after (right column) and before (left column) super-resolution

TABLE 1. The number of characters recognized in the eight plates (Figure 12) before and after super-resolution using the proposed method

License-plates (LP)	Number of detected characters before SR	Number of detected characters After SR
LP1	4	6
LP2	0	4
LP3	0	3
LP4	0	4
LP5	0	8
LP6	0	2
LP7	0	3
LP8	0	8

For visual assessment, a number of the output images before and after super-resolution have been shown in Figure 16.

5. CONCLUSION

A multi-frame super-resolution method using the interpolation technique was proposed. In this method first, we register all frames of video on the reference frame. Then the plate regions are cropped from the frames. The plate region of all frames is then registered to the plate on the reference frame. For each pixel of the up-sampled plate image, a list of the nearest neighbors from all frames is obtained. Using the weighted interpolation, a super-resolution image of the license plate is constructed. Experimental results showed significant improvement in SSIM criteria and license plate recognition. After performing Super Resolution on eight Iranian license plates, Satpa Plate recognition software obtains 53.12% more accuracy in character recognition.

6. REFERENCES

1.   Mjolsness, E., "Fingerprint hallucination", Ph.D. dissertation, California Institute of Technology, (1985).

2.   Tsai, R., "Multiframe image restoration and registration", *Advance Computer Visual and Image Processing*, Vol. 1, (1984), 317–339.

3. Nasrollahi, K. and Moeslund, T. B., "Super-resolution: a comprehensive survey", *Machine Vision and Applications*, Vol. 25, No. 6, (2014), 1423–1468.

4. Irani, M. and Peleg, S., "Super resolution from image sequences", In Proceedings of 10th International Conference on Pattern Recognition (Vol. 2), IEEE, (1990), 115–120.

5. Stark, H. and Oskoui, P., "High-resolution image recovery from image-plane arrays, using convex projections", *Journal of the Optical Society of America A*, Vol. 6, No. 11, (1989), 1715–1726.

6. Amiri, M., Ahmadyfard, A., and Abolghasemi, V., "A fast video super resolution for facial image", *Signal Processing: Image Communication*, Vol. 70, (2019), 259–270.

7. Cheeseman, P., Kanefsky, B., Kraft, R., Stutz, J. and Hanson, R., "Super-resolved surface reconstruction from multiple images", In Maximum entropy and bayesian methods, Springer, Dordrecht, (1996), 293–308.

8. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B. and Salesin, D. H., "Image analogies", In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, (2001), 327–340.

9. Seibel, H., Goldenstein, S. and Rocha, A., "Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos", *IEEE Access*, Vol. 5, (2017), 20020–20035.

10. Mallikarachchi, L. and Dharmaratne, A., "Super Resolution to Identify License Plate Numbers in Low Resolution Videos", In Proceedings of the 7th International Symposium on Visual Information Communication and Interaction, (2014), 220–223.

11. Lee, Y., Jun, J., Hong, Y. and Jeon, M., "Practical License Plate Recognition in Unconstrained Surveillance Systems with Adversarial Super-Resolution", (2019). arXiv preprint arXiv:1910.04324.

12. Vasek, V., Franc, V. and Urban, M., "License Plate Recognition and Super-resolution from Low-Resolution Videos by Convolutional Neural Networks", In Proceeding of British Machine Vision Conference, (2018), 1–12.

13. Zou, Y., Wang, Y., Guan, W. and Wang, W., "Semantic Super-resolution for Extremely Low-resolution Vehicle License Plate", In ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, (2019), 3772–3776.

14. Ojansivu, V. and Heikkila, J., "Image registration using blur-invariant phase correlation", *IEEE Signal Processing Letters*, Vol. 14, No. 7, (2007), 449–452.

15. Seibel, H., Goldenstein, S. and Rocha, A., "Fast and effective geometric k-nearest neighbors multi-frame super-resolution", In 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE, (2015), 103–110.

16. SHAHAAB, Image Processing and Deep Learning Solutions, [Online] https://shahaab-co.ir/license-plate-recognition-library/.

Persian Abstract

چکیده

تشخیص نویسه‌های پلاک خودرو، توسط نرم افزارهای پلاک‌خوان در شرایطی که تصاویر، تفکیک پذیری پایینی دارند، به صورت کامل امکان پذیر نیست. استفاده از دوربین‌هایی که تصاویر با تفکیک پذیری بالا ثبت می‌کنند و برای کار تشخیص پلاک طراحی شده‌اند یک راه حل گران برای حل این مشکل است. اما روش‌های فراتفکیک سازی تصویر می‌تواند یک راه حل ارزان به ما ارائه دهد. ما در این مقاله سعی می‌کنیم با استفاده از رشته تصاویر ویدئویی که از یک پلاک ثبت شده اند تفکیک پذیری پلاک خودرو را توسط یک روش درونیابی هوشمند، بالا ببریم و یک ورودی مناسب برای نرم افزارهای پلاک خوان تهیه کنیم. در این روش، میزان دقیق جابجایی، چرخش و نسبت بزرگ نمایی هر رشته تصویر نسبت به رشته تصویر مرجع محاسبه شده و به کمک اینها درونیابی صورت گرفته و تصویر با تفکیک پذیری بالا، ایجاد می‌شود. در این تحقیق از تصاویری استفاده شده که نویسه‌ها به هیچ عنوان قابل شناسایی توسط پلاک خوان مورد استفاده نبود و با استفاده از روش پیشنهادی برای بعضی تصاویر همه نویسه‌ها و برای برخی چند نویسه پلاک قابل شناسایی شدند.