Method for Improving an Image Segment in a Video Stream Using Median Filtering and Blind Deconvolution Based on Evolutionary Algorithms[*]

Andrey Trubakov[0000-0003-0058-1215] and Anna Trubakova[0000-0003-0280-1760]

Bryansk State Technical University, Bryansk, Russia
trubakovao@gmail.com, trubakovaaa@gmail.com

Abstract. Video surveillance systems, dash cameras and security systems have become an inescapable part of the most institutions ground environment. Their main purpose is to prevent incidents and to analyze the situation in case of extemporaneous events. Though as often as not it is necessary to increase an image segment many times over to investigate some incidents. Sometimes it is dozens of times. However, the obtained material is mostly of poor quality. This is connected either with noise or resolution characteristics, including focal distance. The paper considers an approach for improving image segments, which were obtained after multiple zooming. The main idea of the proposed solution is to use methods of blind deconvolution. In this case, the selection of restoration parameters is carried out using evolutionary algorithms with automatic evaluation of the result. That seems like the most important detail here is pre-processing besides noise minimization within the image, because when the image is repeatedly enlarged the effect of the noise component also increases. To avoid this thing, we suggest using ordinal statistics and average convolution for a series of images. The proposed solution was implemented as a software product, and its operation was tested on a number of video segments made under different shooting conditions. The results are presented at the end of this article.

Keywords: Image Processing, Recovery Images, Super Resolution, Averaging Convolution, Evolutionary Algorithms.

1 Introduction

High-quality image improvement at multiple zooming (super-resolution, SR) has always been and still remains an actual and popular area of mathematical and algorithmic software development. The popularity of these algorithms is due to the fact that no matter how large the resolution of modern cameras is, it is still insufficient and will always be lacking [1,2]. For example, video surveillance cameras are now widespread.

However, in practice, it turns out that the road hog was at a sufficiently large distance and despite repeated zooming, he becomes unrecognizable, the car license plate is not readable, etc. In this case, either simple improving of images or restoring distortions on them are not enough [3], as it is essential to use methods, which allow new data obtaining, removing distortions and improving the frame. In this regard, there is a great need to develop such methods and algorithms that can work with images being zoomed dozens of times. This article is devoted to this problem and the development of such methods.

2   Image segment restoration after multiple zooming

One of the multiple zooming problems is the distortion of the image segment, its susceptibility to a large amount of noise, blooming and blurring, which is not visually perceived at all and is often not suitable for analysis without additional processing. Therefore, it is very important to return as much information as possible and restore the image.

Currently, there are three full-fledged directions of healthy development in the field of super-resolution, which differ in the principles of operation and master data, entering the system input. The first direction is simple image zooming with post-processing of the resulted image. A single frame is used as the source images. At the same time in case of simple zooming by means of the nearest neighbor method the image will look like a mosaic made up of squares when being zoomed repeatedly, which is bad enough. Therefore, the approaches in this category are aimed at removing the mosaic effect, making the image more readable and beautiful without adding new details to it. The most well-known representative of this class of algorithms is interpolation-based approaches (for example, the bicubic interpolation method [4,5]). In the second class of algorithms, an attempt is made either to obtain new details by repeatedly zooming or improving the visual perception of an image based on a single snapshot. Very often these methods are based on self-learning systems with a set of samples. It is a preselected training sample that makes adding of details, which were out of situation in the process of manifold zooming, possible. In this case, the system makes a sort of predictions and assumptions about what might take place in a particular place inside the image. Among the most popular approaches in this category are neural networks, random Markov fields, and others [6-9].

The third class of algorithms is aimed at working with a series of images. In this case, it is assumed that there is not just a separate image, but a video stream, a series of snapshots of the same object. This allows trying to combine data from different images, extract different details that are possessed by one image but be lacking in another frame and also reducing the impact of noise and other distortions. A number of retrain positive frames provides more opportunities, but also requires the use of more complex approaches and algorithms. This work is devoted to the development of the third group of methods. It offers an approach of truncated convolution of a series of images taking into account ordinal statistics, which allows reducing both: the influence of the noise component and distortions in the averaging process, thereby increasing the influence of

the sections having maximum information content. After that, we suggest using the method of blind deconvolution based on evolution algorithms to diminish defocusing. The scheme of this option is shown in Fig. 1.
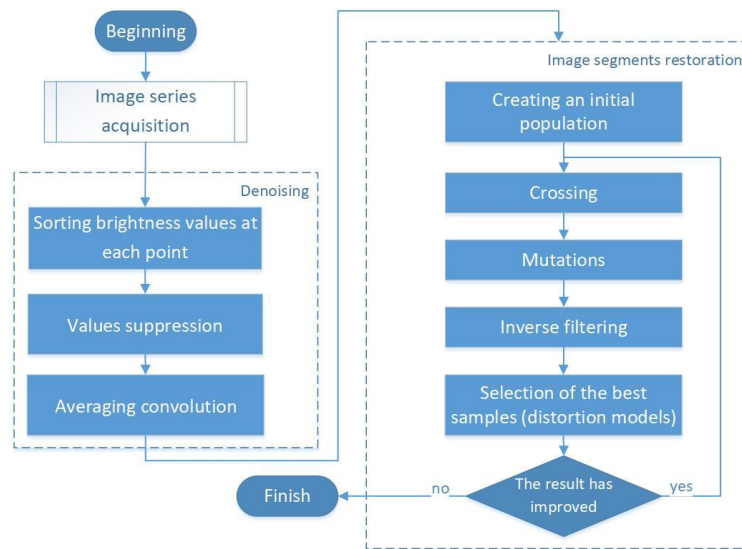

Fig. 1. Diagram of the restoration algorithm and a segment of zooming.

As can be seen from the scheme presented above, the proposed algorithm has two stages of interest - reducing the influence of noise using information from a series of images and blind deconvolution based on evolutionary algorithms. Let's look at these steps to some detail.

3   Using ordinal statistics and convolution to increase image quality

Multiple zooming makes you solve a number of problems. And one of them is the influence of the noise component. Any image received from a camera or a still-shot camera has interference and noise. Noise may not be visible to the naked eye at normal resolution, but it can be highly visual if enlarged. This is due to the fact that when we make zooming, we increase the contribution of noise to the result owing to the expansion in the amount of it.

Noise also greatly affects the result if the restoration process is used further (for example, blind deconvolution based on the Wiener filter). Most restoration methods are very sensitive to noise and can result in the wrong way. If you have a whole series of synchronized images (a video stream) instead of a single snapshot, you can use ordinal statistics and a truncated average as preprocessing. The median filter was considered as an alternative, but it showed a slightly worse result in the presence of combined noise. You can use a winsorized mean filter for these purposes also.

Let's look at this approach in more detail. Let's assume that there is an array of images of the same object: its shooting conditions did not change during the entire series:

$$\{Img_1, Img_2, \ldots, Img_n\}. \tag{1}$$

In ideal shooting conditions, the following condition must be met: the value of any pixel in the first image in the position [i, j] must completely match that of all other images in the series:

$$\forall\, i, j, k\colon Img_1[\mathrm{i}, \mathrm{j}] == Img_k[\mathrm{i}, \mathrm{j}]. \tag{2}$$

However, this is not the case in practice. A number of images have a noise component. We will assume that the noise is a random variable (in most cases, it is). In this case, the $Img_k[i,j]$ pixel will have a real value in a number of images, and a distorted value for some of them. In order to reduce the effect of distortion, you can find the arithmetic mean of each pixel within the entire series consisting of *n* images (we choosen *n* equal to 20 in our experiments):

$$Img_k[\mathrm{i}, \mathrm{j}] = \sum_{i=1}^{n} \frac{Img_k[\mathrm{i,j}]}{n}. \tag{3}$$

The averaging results in the noise component as if got smeared across the entire series of images, thereby reducing the contribution to each specific image. However, experiments have shown that it is better to use a truncated average instead of simple averaging. That is why first it is critical to sort the values of each specific pixel by brightness value. Let's assume that the interference caused either a slight increase in brightness at this point or inversely, its decrease. So it is necessary to discard the first and last m statistics, and average the remaining part:

$$Img_k[\mathrm{i}, \mathrm{j}] = \sum_{i=m}^{n-m} \frac{Img_k[\mathrm{i,j}]}{n}. \tag{4}$$

As shown by experiments on real images, this preprocessing gives more accurate results during further restoration and an increase in the convergence rate of the restoration process by approximately 10%, when based on the evolutionary algorithms proposed and described later in this article.

4   Blind deconvolution based on evolutionary algorithms

The next step, after receiving a snapshot with a reduced impact of the noise component, is the restoration stage, shown on the right in the General diagram (Fig. 1). As proved by numerous experiments, after multiple zooming, the segment of interest is almost always blurred and poorly readable. This is comparable to an image taken with the wrong focal length. Therefore, the logical step here is the attempt to use methods for restoring distorted images based on inverse filtering.

Conventional methods of inverse filtering allow restoring the image quite well. But these methods have one great disadvantage – it is inevitable to give a detailed description of the distortion model, including all its parameters, such as blur radius and noise.

If you choose one of the parameters incorrectly, the restoration result will be worse than even the original image, and the restoration will nothing but spoil the image. With multiple zooming, it is not possible to predict the distortion parameters and build a full-fledged model of it.

To solve this problem, methods of blind reverse convolution (blind deconvolution) are actively used. The purpose of this approach is to find an unknown distortion function and its parameters. To speed up selection, we suggest using genetic algorithms, and to evaluate the result at the stage of choosing alternatives, we use automatic metrics based on gradient and maximum likelihood estimates.

The use of genetic algorithms cannot be attributed to traditional approaches for image processing, but in some cases they work giving not bad results [10]. To use the genetic algorithm it was necessary to define and formalize the following concepts:

- biological units of population (distortion model variants);
- principles of crossing;
- principles of mutations;
- function of determining biological unit adaptability (assess alternatives).

Biological units of population in genetic algorithm. The task of blind deconvolution is to find a distortion model in which the restored image will have the highest quality. Therefore, the problem is reduced to finding the maximum of a certain evaluation function, the input parameters of which serve as the parameters of the distortion model. Based on the above assumption, a blur model with a certain circle radius was chosen as the distortion model. The selection of the model type was by trial and error. This model showed better results than other distortion models (blur, motion blur with Gauss function, and curvilinear distortion models were analyzed as exponents). The parameters of the distortion model in this case are the blur radius and the signal-to-noise ratio. The Wiener filter was used as an inverse filter to restore the image. It has given a good account of itself in systems for restoring distorted images. Units of the initial population for use in the genetic algorithm represent various variants of the blurring radii of the distortion model. The starting population contains samples with values covering the range of possible parameter values in the best possible way. When implementing the system, it was decided to choose uniform values in the range from 0 to 10% of the length of the largest side of the image segment being restored as the blur radius. This choice gives the best performance of the genetic algorithm, less often converges to the local maximum, and provides fairly good indicators for the time of operation.

However, later, in the course of conducting experiments, it was observed that the choice of the initial options for the blur radius made formalized dependency on the zoom scale more difficult. The smaller the zooming, the lower the maximum blur value can be selected in the original population. But in some cases, when the original image is multiply zoomed, 10% is not enough. However, this fact is not yet fully developed and requires further research and analysis in the future.

Crossing and mutations. The main stages in the work of the genetic algorithm are crossing and mutation. When crossing different biological units, the binary exchange of parts of information encoding the radius of blurring occurs. In this way, the model parameters are modified and a new population is formed for evaluation.

Right after crossing, a number of exponents undergo mutation. The paper identifies two types of mutations: weak and strong. In the case of weak mutations, the blur radius changes to a random value of a small nominal value (up to 3, the value was chosen a posteriory for increasing the convergence rate). Such mutations lead to the modification of units and allow expanding beyond the original population in the case when the population includes samples with very similar blurring radii. Areas with mutations are also added to the resulting population.

The second type of mutation (strong mutations) is performed with a significant change in the blur radius. For this type of mutation, only a few biological units are selected and the blur radius in them changes by a random amount up to 5% of the image size. This stage allows exiting of the local maxima areas. However, a large number of such mutations just contributes to the delays of the convergence process. Therefore, the number of such mutations was selected in the region of 2 per iterate.

Evaluating function of the image restoration quality. After obtaining the population (variants of the distortion model), they must be evaluated using the goal function to select the most adapted samples (images with good restoration quality). The goal function was chosen to reverse restoration using a Wiener filter and to estimate the average gradient value of the resulting image. The gradient proved to be quite good when restoring defocused images. The Sobel operator was used to calculate the gradient.

The choice of gradient as an evaluation function was made for the following reasons. Research in the field of image processing and analysis shows that despite the variety of shapes of the images themselves, their gradient histogram is of similar appearance [10,11]. The gradient distribution has a peak in the region of zero and decreases to the maximum value. Moreover, the decrease is not according to the Gauss distribution, but has considerably larger values in the region of high values. An example of gradient distribution is shown in figure 2.
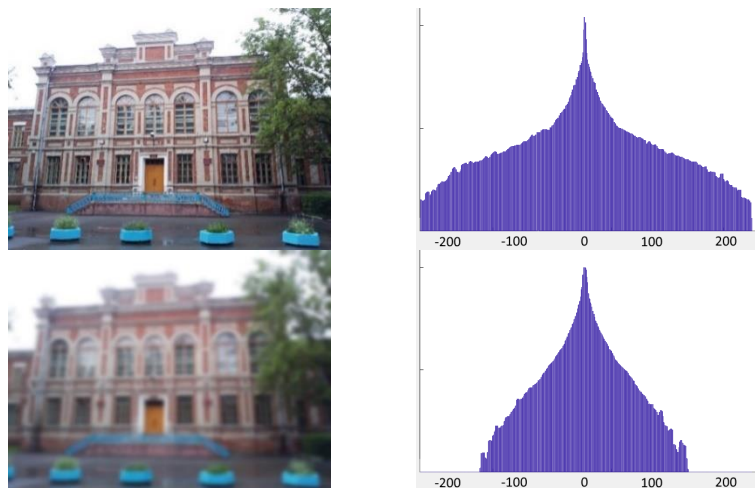

Fig. 2. Gradient distribution for sharp and blurry images (the photo shows Bryansk state technical University).

This agrees with the intuitive notion to a great extent. Most images have large areas with more or less constant brightness (where the gradient value is close to zero) and the borders of these regions-objects usually end with quite sharp differences in brightness, which is up to a large value of the gradient. As a result, we have a rather peculiar but stable distribution of the gradient, which is shown in Fig. 2. Thus it can be seen that the more blurred is the image, the fewer high values of the gradient it has. The form of distribution remains the same, but the image becomes narrower.

This estimation approach is very popular among blind deconvolution methods. However, in the course of experiments, it was noticed that it is not suitable for images with a large number of sharp changes (for example, the image of text on a white background). In addition, it has one more disadvantage: for evaluation of the image quality the etalon gradient histogram is required to be compared with, which is not always convenient. In real conditions, it is not always possible to select such a histogram in advance, and the use of some average analogue often does not exceed the use of a simple value of the average gradient over the entire image. Therefore alternatively in our system, we used a slightly different approach, which also uses the gradient value, but focuses on the average value across the entire image. In this case, you do not need to set the etalon gradient histogram, and the system adapts itself to the current frame and the data type in it.

Additional restrictions of the goal function. Using the average gradient generally gives very good results. However, sometimes there are situations in which a high gradient value does not produce better results. This can happen with images having a large number of small objects (Fig. 3a), which turn into dots when blurred, thereby increasing the average gradient value (an example of this case is shown in figure 3b).

Fig. 3. An example of restoration based on the average gradient of a photo of a commemorative inscription on a stand in Lenin Public Gardens, Bryansk: a) the original segment, b) restoration by maximizing the average gradient, c) the best option, selected manually.

To eliminate the defect additional restrictions when evaluating the quality of alternatives were set. The MSE (Mean Square Error) metric was chosen as the restriction,

which is often used to evaluate the similarity of the restored image with the distorted one. A large value of this metric shows a significant difference between images. Using this metric, as can be seen the images in Fig. 3a and 3b are within only 6 conventional units (which is inappreciable). So these alternatives should not be considered and a number of them should be excluded from the resulting sample (but not all of them, in order not to exclude absolutely the entire population providing that the restored image really should be similar to the original one).

Optimization completion criteria. As a completion criteria of restoration stage a number of populations was chosen, which went wrongly in regard to the restoration result. Ten populations were selected acting as this value. After completing the iterative algorithm, the value of one of the biological units is selected as parameters of the distortion model and used to restore the image. The result is shown in figure 3c.

Nevertheless, as experiments have shown, it is not always necessary to choose a sample with the best indicator of the goal function. This is due to the fact that the average gradient estimate does not always give an optimal result if the population units are close in value to the goal function. Therefore, the following algorithm was proposed to select the final version of the distortion model.

1. Select two best samples from the point of view of the goal function.
2. If the value of the goal function of the first sample significantly exceeds the value of the second one (the difference in the values of the average gradient is more than 15), then the first is selected as the most suitable. This is a situation where there is an obvious favorite and it has not changed for a long time (more than 10 iterations).
3. If the value of the goal function of the first and second units do not differ much (the difference is less than 15, samples are similar to each other), then the structure similarity metric (SSIM) is calculated for the restored images to compare with the original one. As a result of comparison, the image with less SSIM metric value is selected (since it is necessary to select an image resembling a distorted version to a lesser extent).

Applying another metric (SSIM instead of MSE used in the previous stage) showed a better result. The SSIM metric did not perform in the best way (although it was quite good) at the intermediate stage of evaluation, while at the final stage it allowed achieving a qualitative result.

5   Observation

For experimental study a utility was developed for processing data from a video stream with the ability to improve separate segments. The utility was written in C++ using the OpenCV library.

A number of videos of the sights of the city of Bryansk were taken as source data. Shooting was done on a camera with a tripod. This is due to the fact that this work does not concern issues related to the mapping and synchronization of frames during processing. Frame synchronization is a feature length topic that requires separate study. In

addition, shooting from a tripod is the most approximate for the conditions of use in surveillance cameras, in which the camera itself is fixed rigidly.

A series of experiments was performed, the purpose of which was to search for and identify the strengths and weaknesses of the proposed method, determine the limits of its applicability, and confirm its performance.

The first series of experiments was aimed at determining the effectiveness of the first stage of the system – filtering a series of images based on ordinal statistics and average convolution. An example of this is shown in figure 4.

Fig. 4. Example of averaging filter working in a series of images.

The figure gives an example of one of the video stream frames, which shows the Bryansk regional Puppet Theater. With sufficient zooming of the image segment, being highlighted by the marker, you can see some noise that hamper perception and obstruct the view (enlarged segment is shown on the left). But a detailed analysis of a series of images proves that these interferences are random. Therefore, the proposed version of preprocessing based on ordinal statistics and average convolution really produces a successful result (the segment after processing is shown on the right).

However, it is worth noting some minor effect of the proposed filter. The image becomes slightly out of focus and a number of objects will increase slightly in linear dimensions. This effect is most likely due to incomplete synchronization of images during averaging. This side effect can be eliminated by supplying an additional synchronization stage to the filter based on calculating key points. Though, at the moment we have not conducted such studies yet and this is the direction of further development of the work.

The second series of experiments was to test the efficiency of the distortion restoration stage based on blind deconvolution using a genetic algorithm. An example of restoring of two image segments is shown in figure.
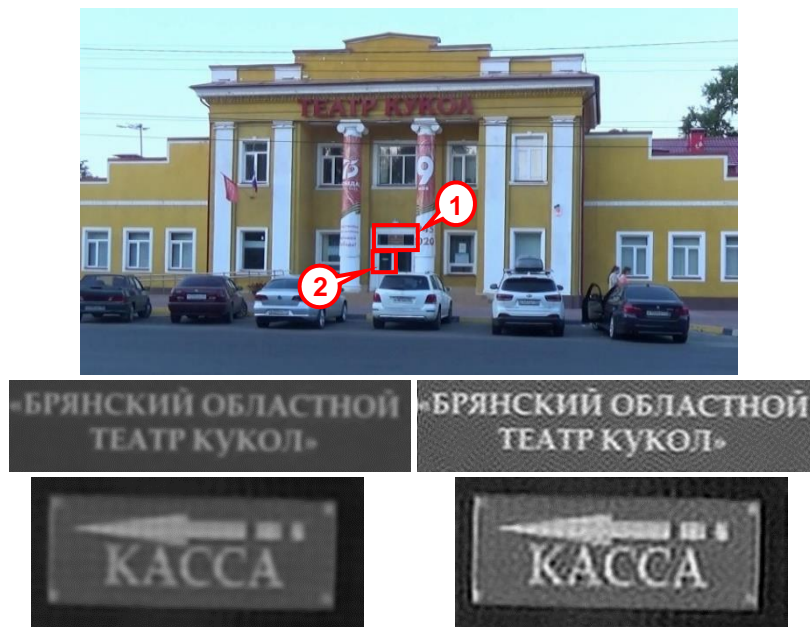

Fig. 5. Examples of image segment restoration stage.

As an example of restoration, we selected two tablets in the image of the Bryansk regional Puppet Theater, blurred after enlargement due to the fact that they fell out of the camera's focal length (the enlarged segments are shown in the figure on the left). After the blind deconvolution stage, this defocus was significantly reduced and the outlines of the letterings became clearer (the result is shown in the figure on the right).

Nevertheless, it is worth making a number of observations. First, like any method of blind deconvolution, this approach can lead to side defects (for example, ripples or false contours). This effect is not always apparent and in some cases is almost invisible and does not affect further analysis. But if such behavior is undesirable, it can be avoided by making parameter adjustments at the last stage.

Secondly, the described method cannot be considered a panacea in absolutely all cases. As shown by the research, it does not give bad results if zoomed and processed image segment was not in the focus of the lens (which is a fairly common situation when analyzing frames from video surveillance systems that require additional processing). But it does not enable you to restore image details if there is no information about them at all in the video stream (for example, when increased tenfold, if the analyzed segment is up to several pixels in the original image).

6   Conclusion

The paper presents a method for improving image segments obtained from a video stream. The main features of this method are using of truncated average convolution at the preprocessing stage and applying of blind deconvolution based on genetic algorithms. Research has shown that in some cases this approach is quite effective, especially in a situation where the enlarged image segment is not in the focal length of the camera. However, the proposed solution can be developed more in the direction of further studies.

References

1. Wenming, Y., Zhang, X., Tian, Y., Wang, W., Xue, J.-H.: Deep Learning for Single Image Super-Resolution: A Brief Review (2019).
2. Dai, S., Han, M., Xu, W., Wu, Y., Gong, Y., Katsaggelos, A.K.: Softcuts: a soft edge smoothness prior for color image super-resolution. IEEE Transactions on Image Processing, vol. 18, no. 5, pp. 176-179 (2009).
3. Trubakov, A.O., Prazdnikova, T.D.: Restoration of distorted image areas. Proceedings of the 28th International Conference on Computer Graphics and Vision, Moscow, P.300-303 (2018).
4. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview, IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36 (2003).
5. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 6, pp. 1153-1160.
6. Cao, F., Cai, M., Tan, Y., Zhao, J.: Image super-resolution via adaptive regularization and sparse representation. IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 7, pp. 1550-1561 (2016).
7. Liu, J., Yang, W., Zhang, X., Guo, Z.: Retrieval compensated group structured sparsity for image super-resolution. IEEE Transactions on Multimedia, vol. 19, no. 2, pp. 302-316 (2017).
8. Zhang, K., Tao, D., Gao, X., Li, X., Li, J.: Coarse-to-fine learning for single-image super-resolution. IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 5, pp. 1109-1122 (2017).
9. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. Proceedings of the European Conference on Computer Vision, pp. 184-199 (2014).
10. Savostin, I.A., Trubakov, A.O. The Combination of Morphological and Evolutionary Algorithms for Graph-based Pathfinding on Maps with Complex Topologies. Proceedings of the 29th International Conference on Computer Graphics and Vision, Moscow, pp.300-303 (2019).
11. Sun, J., Xu, Z., Shum, H.-Y.: Image super-resolution using gradient profile prior. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8 (2008).
12. Yan, Q., Xu, Y., Yang, X., Nguyen, T.Q.: Single image superresolution based on gradient profile sharpness. IEEE Transactions on Image Processing, vol. 24, no. 10, pp. 3187-3202, (2015).