The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric

Frédérique Crété-Roffet, Thierry Dolmiere, Patricia Ladret, Marina Nicolas

The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric

Frederique Crete[a,b], Thierry Dolmiere[a], Patricia Ladret[a], Marina Nicolas[b]

[a]Laboratoire des Images et des Signaux,
46 avenue Felix Viallet, 38031 Grenoble, France;
[b] STMicroelectronics,
12 rue Jules Horowitz, 38019 Grenoble, France.

ABSTRACT

To achieve the best image quality, noise and artifacts are generally removed at the cost of a loss of details generating the blur effect. To control and quantify the emergence of the blur effect, blur metrics have already been proposed in the literature. By associating the blur effect with the edge spreading, these metrics are sensitive not only to the threshold choice to classify the edge, but also to the presence of noise which can mislead the edge detection.

Based on the observation that we have difficulties to perceive differences between a blurred image and the same re-blurred image, we propose a new approach which is not based on transient characteristics but on the discrimination between different levels of blur perceptible on the same picture.

Using subjective tests and psychophysics functions, we validate our blur perception theory for a set of pictures which are naturally unsharp or more or less blurred through one or two-dimensional low-pass filters. Those tests show the robustness and the ability of the metric to evaluate not only the blur introduced by a restoration processing but also focal blur or motion blur. Requiring no reference and a low cost implementation, this new perceptual blur metric is applicable in a large domain from a simple metric to a means to fine-tune artifacts corrections.

Keywords: Blur, Perception, No-Reference, Objective metric, Subjective test.

1. INTRODUCTION

New technologies of displays, panels, cameras or mobiles aim at obtaining the best image quality. This phenomenon plays an important role in the evolution of the image correction algorithms to remove noise or compression artifacts. Unfortunately, most of the corrections using low-pass filters not only smooth artifacts, but also lose a part of the high frequency content and generate a blur effect. Others methods such as rescaling algorithms can also generate this artifact.

To control and quantify the emergence of the blur effect, blur metrics have been proposed in the literature. One of the first blur estimation [1] was established in 1997 to improve the super resolution algorithms. Then, the blur estimation was used as a quality metric. In 1999, Marichal et al [2] estimated the blur by using an histogram of the DCT coefficients, but this method was limited to the frames in the compressed domain. Then, several metrics based on an edge detection were proposed. Among them, Marziliano et al [3] took into account the edge spreading using the inflection points delimiting an edge. Caviedes et al [4] made an evaluation of the sharpness (the inverse of the blur) by computing the Kurtosis of the DCT on blocks containing edges. Ong et al [5] characterized the amount of blur by computing the average extent of the edges. Based on a given edge detector, these metrics are sensitive not only to the threshold choice to classify the edge, but also to the presence of noise which can mislead the edge detection. Hu et al [6] proposed a Gaussian blur estimation algorithm which is not based on an edge detection. Their method model the focal blur with the normalized Gaussian function and is well adapted for the out of focus blur detection in images or videos.

To be independent from any edge detector and to be able to predict any type of blur annoyance, we propose a new approach which is not based on transient characteristics but on the discrimination between different levels of blur perceptible on the same picture. In fact, we observe that we have difficulties to perceive differences between a blurred image and the same re-blurred image. Consequently, we use this phenomenon to estimate the blur annoyance.

The major interest of a no-reference metric is to be able to replace subjective tests requiring time and means. Nevertheless, these tests are necessary to validate the metric by correlating them with the human judgment. Following the ITU recommendations [7] to make subjective tests, we take also advantage of this step to analyse the behaviour of the viewers and their abilities to evaluate the blur effect on a picture.

In this paper, we present the description and the validation of our no-reference perceptual blur metric. Section 2 presents the method used to evaluate without reference the blur perception, section 3 is the description of the experimental protocol used to establish subjective tests, and section 4 proposes an analysis of the behaviour of the viewers and shows the correlation between the objective measurements of the quality metric and the subjective tests. Then, section 5 lists the possible applications of the metric and section 6 provides the conclusion of the paper

2. THE NO-REFERENCE BLUR ESTIMATION

2.1 The blur discrimination of the human perception

Because the blur effect is caused by a loss of the high frequency content, it can be reproduced with a low-pass filter. By studying naturally blurred pictures and using different low-pass filters to cover the largest possible range of blur levels, we observe that we have difficulties to perceive differences between a blurred image and the same re-blurred image. In Figure 1, we present from left to right the original sharp picture, the original picture blurred with a low-pass filter and the blurred picture re-blurred with the same low-pass filter. We observe a high difference in term of loss of details between the first and the second picture and a slight difference between the second and the third picture.

Figure 1: from left to right: original picture, original picture blurred with an averaging filter, blurred picture re-blurred with the same averaging filter.

Actually, the more we blur, the more the neighboring pixels converge to the same gray level. If we blur a sharp picture, gray levels of neighboring pixels will change with a major variation. On the contrary, if we blur an already blurred picture, gray levels of neighboring pixels will still change, but only to a weak extent. We can explain this phenomenon by the fact that the second blur effect reduces a difference between pixels that has already been reduced by the first blurring effect. To illustrate this phenomenon, the Figure 2 represents the three histograms of the absolute difference between neighbouring pixels for a same picture sharp, blurred and re-blurred. We notice that the high differences significantly decrease after the first blurring step and slightly decrease after the second blurring step. On the contrary, the number of slight variations increases with the two blurring steps.

The key idea of our blur estimation principle is to blur the initial image and to analyse the behaviour of the neighbouring pixels variation. The blurring step should be done with a strong low-pass filter in order to be sure to compare the initial image with an image which seems blurred for the human perception. The choice of the type of filter is not exhaustive if it is a strong filter.
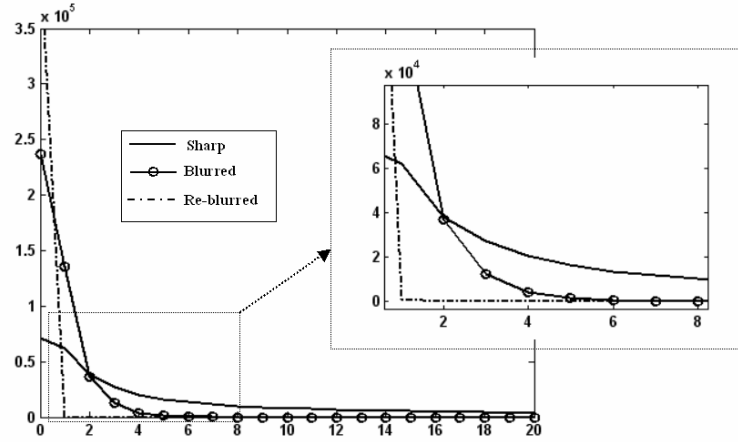
Figure 2: histograms of the variations (the absolute differences) between neighboring pixels for a sharp, blurred and re-blurred picture

2.2 General observations on the blur perception

In addition to this previous principle, we note others phenomena on the blur perception which lead us to a specific principle of the blur estimation.

The first observation depends on the local content of the picture. If we look at a picture containing a small blurred part over a homogeneous area, we perceive it as an non-sharp picture even if only a minor part of this picture is blurred (Figure 3). For this reason, in the analysis of the neighboring pixels variation, we only take into account the pixels which have changed after the blurring step.

Figure 3: A small blurred text over a large flat area.

Moreover, by blurring a sharp edge on a flat area, we observe that we create slight variations between neighboring pixels around the edge which were not existing on the sharp picture. To avoid to take into account these variations, we consider only the neighboring pixels variations which have decreased after the blurring step. This method takes advantage of the possibility to access to specific local variations representatives of the blur effect.

Finally, by using our principle, we are robust to an eventual presence of noise. In fact, more or less additive noise makes blurred pictures appear more or less sharp. By taking into account all variations of the image which have decrease after the blurring step, we analyze also the pixels containing the noise information which can be located on the edges, on the textured areas or on the homogeneous areas.

The last observation is done for pictures with a motion blur. For example, we observe this phenomenon for a picture extract from a traveling sequence. On a picture blurred in one direction, the eye perceives the blur effect on this direction but does not perceive the sharpness on the others directions. For this reason, we estimate the blur in the horizontal and the vertical directions and select the blur the more annoying as the final blur value.

2.3 Description of the blur estimation principle

Based on the phenomena explained in the previous section, we are able to quantify the blur annoyance on a picture by blurring it and comparing the variations between neighboring pixels before and after the low-pass filtering step. Consequently, the first step consists in the computation of the intensity variations between neighboring pixels of the input image. On this same image, we apply a low-pass filter and compute also the variations between the neighboring pixels. Then, the comparison between these intensity variations allows us to evaluate the blur annoyance. Thus, an high variation between the original and the blurred image means that the original image was sharp whereas a slight variation between the original and the blurred image means that the original image was already blurred. This description is summarized in Figure 4.
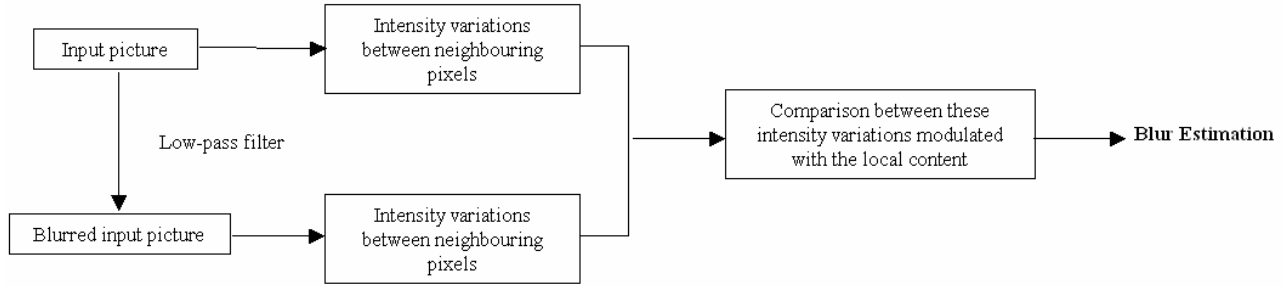

Figure 4: Simplified flow-chart of the blur estimation principle

Based on this principle, we describe in the following subsection the algorithm of this no-reference blur metric.

2.4 Algorithm description of the no-reference blur metric

It is known that the sharpness of an image is contained in its gray component. This assumption, which is verified with subjective tests, justify that we estimate the blur annoyance only on the luminance component. The flow chart in Figure 5 describes the steps of the algorithm description and refers to the following equations.

Let F be the luminance component of an image or a video frame of size of $m \times n$ pixels. To estimate the blur annoyance of F the first step consists in blurred it in order to obtain a blurred image B. We choose an horizontal and a vertical strong low-pass filter (1) to model the blur effect and to create B_{Ver} and B_{Hor}.

$$h_v = \frac{1}{9} \times [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1] \qquad h_h = transpose(h_v) = h_v' \tag{1}$$

$$B_{Ver} = h_v * F \qquad\qquad B_{Hor} = h_h * F$$

Then, in order to study the variations of the neighboring pixels, we compute the absolute difference images D_F_{Ver}, D_F_{Hor}, D_B_{Ver} and D_B_{Hor} as followed:

$$D\_F_{Ver}(i,j) = Abs(F(i,j) - F(i-1,j)) \quad for\ i=1\ to\ m-1,\ j=0\ to\ n-1$$
$$D\_F_{Hor}(i,j) = Abs(F(i,j) - F(i,j-1)) \quad for\ j=1\ to\ n-1,\ i=0\ to\ m-1$$
$$D\_B_{Ver}(i,j) = Abs(B_{Ver}(i,j) - B_{Ver}(i-1,j)) \quad for\ i=1\ to\ m-1,\ j=0\ to\ n-1$$
$$D\_B_{Hor}(i,j) = Abs(B_{Hor}(i,j) - B_{Hor}(i,j-1)) \quad for\ j=1\ to\ n-1,\ i=0\ to\ m-1$$

$$\tag{2}$$

As we explain in the previous subsection, we need to analyze the variation of the neighboring pixels after the blurring step. If this variation is high, the initial image or frame was sharp whereas if the variation is slight, the initial image or frame was already blur. This variation is evaluated only on the absolute differences which have decreased:

$$V_{Ver}=Max(0,D\_F_{Ver}(i,j)-D\_B_{Ver}(i,j)) \quad for \ i=1 \ to \ m-1, \ \ j=1 \ to \ n-1$$
$$V_{Hor}=Max(0,D\_F_{Hor}(i,j)-D\_B_{Hor}(i,j)) \quad for \ i=1 \ to \ m-1, \ \ j=1 \ to \ n-1$$

(3)

Then, in order to compare the variations from the initial picture, we compute the sum of the coefficients of D_F_{Ver}, D_F_{Hor}, D_V_{Ver}, D_V_{Hor} as followed:

$$s\_F_{Ver}=\sum_{i,j=1}^{m-1,n-1}D\_F_{Ver}(i,j) \qquad s\_F_{Hor}=\sum_{i,j=1}^{m-1,n-1}D\_F_{Hor}(i,j)$$

$$s\_V_{Ver}=\sum_{i,j=1}^{m-1,n-1}D\_V_{Ver}(i,j) \qquad s\_V_{Hor}=\sum_{i,j=1}^{m-1,n-1}D\_V_{Hor}(i,j)$$

(4)

Finally, we have to normalize the result in a defined range from 0 to 1:

$$b\_F_{Ver}=\frac{s\_F_{Ver}-s\_V_{Ver}}{s\_F_{Ver}} \qquad b\_F_{Hor}=\frac{s\_F_{Hor}-s\_V_{Hor}}{s\_F_{Hor}}$$

(5)

We note that the variations between the two differences images D_F and D_B are always slighter than the values of the initial difference image D_F.

Then, we select the blur the more annoying among the vertical one and the horizontal one as the final blur value.

$$blur\_F=Max(b\_F_{Ver},b\_F_{Hor})$$

(6)

To summarize, we obtain a no-reference perceptual blur estimation $blur\_F$ ranging from 0 to 1 which are respectively the best and the worst quality in term of blur perception.
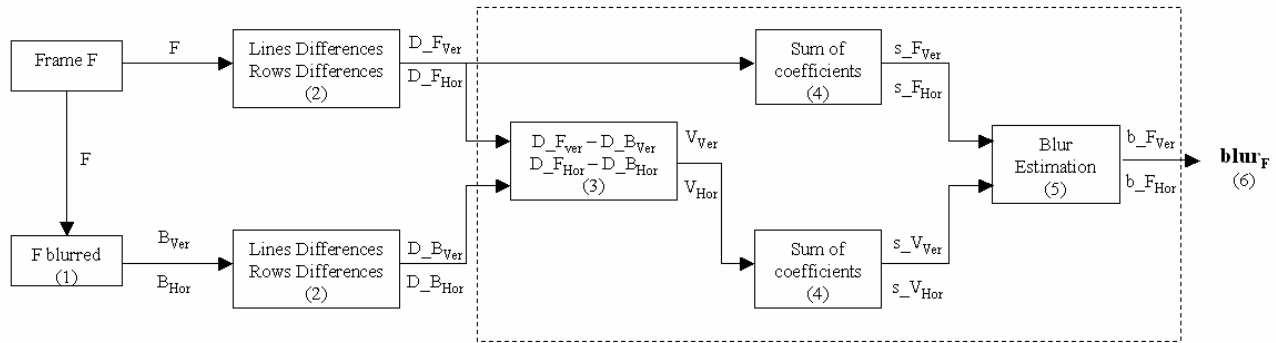

Figure 5: Flow chart of the algorithm with the equations references.

3. EXPERIMENTS

Following the ITU-R recommendations [7], we made one main subjective experiment in order to assess the global quality of some blurred pictures.

3.1 Subjective test procedure

The aim of the procedure is to collect an absolute subjective opinion of each picture. In this way, we develop a program for psychophysics experiments with the psychtoolbox of Matlab. Before this experiment, a training session calibrates the opinion of each viewer with ten pretest pictures. These results are not taken into account in the final data. The experiment is a test-retest experiment, containing two equivalent parts separated by a pause of 3 minutes minimum. Each part contains a whole set of one hundred pictures. For the experiment, we set several timing parameters. Each picture is shown during a duration of 5 seconds. This time has been chosen after many tests to collect only the first spontaneous opinion of the viewers. Nevertheless, the viewers can assess the pictures faster by hitting a button. This allows to obtain results about visualization time. The picture is then replaced by a graph showing a five level graduation. No time limit has been set for the assessment. The viewer can give five scores between 1 and 5 corresponding to several quality levels as shown on the Table 1. After the assessment, 0.5 seconds of a neutral gray is presented before the next picture in order to suppress any retinal effect.

Table1: Five-grade scale

Quality	Score	Impairment
Excellent	5	Imperceptible
Good	4	Perceptible but Not Annoying
Fair	3	Slightly Annoying
Poor	2	Annoying
Bad	1	Very Annoying

3.2 The room

The assessment of the quality of a picture depends on the environment. Thus, we make the main experiment in a total neutral gray environment in order to minimize parasite reflections on screen, walls, ceiling and table. During the experiment, the light was turned off contrary to the pause where the light was on in order to make the viewer's eyes rest (Figure 6).

3.3 The screen

The screen used is a 21" View Sonic G220f which has been previously calibrated with the GretagMacBeth Eye-One in order to have a 6500K color temperature. The resolution of the display is set to a resolution of 1400x1050 in order to fit with the picture's original size. The viewing distance is about 4 times the height of the screen, compliant with the ITU recommendations (Figure 6).

Figure 6: Subjective tests room

3.4 The pictures

The content of the pretest is chosen to represent the different levels of degradation. Then, several filters are randomly applied to 12 pictures: 2 types of Gaussian filters (3x3 and 5x5), 4 types of averaging (3x3, 5x5, 7x7 and 9x9) and motion filters (3x1, 5x1, 7x1 and 9x1) and one average disk filter with a diameter of 11 pixels. The 12 initial pictures represent natural scenes more or less textured. The Figure 7 illustrates a sample of the data set. In each part of the experiment, all pictures are randomly put on the screen and assessed. The random order allows to suppress any influence from the order of the presentation. Finally, 132 pictures are assessed twice per viewer.

Figure 7: Example of pictures from the data base

3.5 The paddle

We only used a paddle to make the experiment. Thus, the viewers are almost twice faster than using a keyboard. Finally, they take about 25 minutes to assess the 264 pictures. The experiment is all the more easy to pass that there are only 3 buttons (validate-confirm-cancel) and 2 arrows to change a score given to a picture.

3.6 The viewers

At the beginning of the experiment, information about the viewers are written in the database (age, sex, activity, ...). An Ishihara test is made in order to detect possible color blindness. Finally, the panel of observers is made up with 15 non-expert viewers. Each one has a perfect visual acuity.

4. RESULTS

A similar experiment with gray scaled pictures and colored pictures has been made before this main one in order to prove that the luminance component study was sufficient to estimate the blur artifact. We notice that the scores corresponding to the gray scaled pictures were different only by 0.21 with a standard deviation of 0.19 on the five-grade scale. That confirms the fact that the analysis of the chrominance is not very important when we evaluate the blur effect.

4.1 Analysis of the viewers behavior

Following the ITU recommendations, the computation of the Mean Opinion Score (MOS) and statistical analysis allowed us to reject 3 viewers out of 15. Then, we examine the different times taken by the viewers to assess a picture, and we obtain some interesting results represented in Figure 8:

- There are no hesitation when the pictures are too blurred.

- On average, the highest scores have been given after a longer time than the lowest scores. Indeed, the viewers wanted to find some defaults before to give the score.

- There is a psychological jump in the subjective evaluation of the blur. Indeed there are less assessments with scores between 2 and 3.5. That implies that the blur is generally perceived or not.
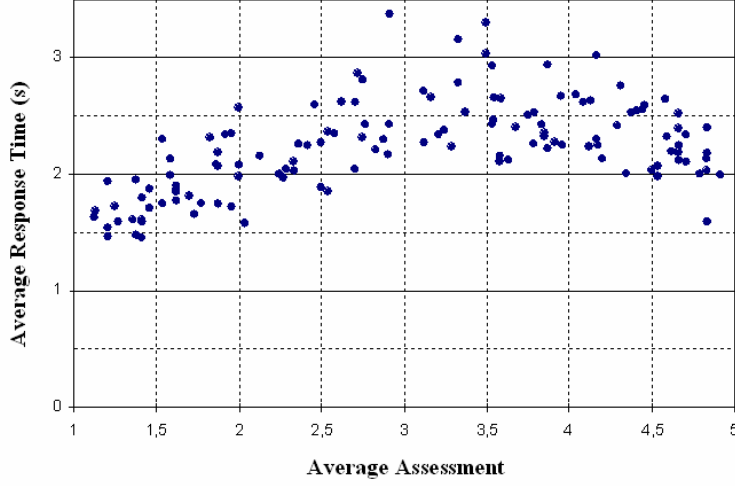
Figure 8: Average response time function of the average assessment

4.2 Correlation between the subjective tests and the no-reference perceptual blur metric

In order to evaluate the proposed algorithm, we compared the subjective and the objective assessments. First, we observe a correlation between the objective and the subjective measurements. The Figure 9, which represents the MOS with the 95 percent interval versus the objective metric illustrates this correlation: the no-reference metric follows the human estimation. In order to evaluate more precisely these results, we try to find a relation between these two assessments. Referring to psychophysics functions, if A_{Obj} is the objective assessment and A_{Sub} the subjective assessment, we tried to approximate the function $ASub = f(Aobj)$ by the following function (7).

$$A_{Sub} = \frac{A}{1+\exp((D*A_{Obj}-D_M)*G)} + B \qquad (7)$$

where the real numbers A and B are found with the boundary values of the function, and G, D and D_M by computing the coefficients of a linear approximation of the following function:

$$\log(I) = (D*A_{Obj}-D_M)*G$$
$$\text{with} \quad I = \frac{1}{p}-1 \quad \text{and} \quad p = \frac{1}{1+\exp((D*A_{Obj}-D_M)*G)} \qquad (8)$$

Finally, we achieve to the equation (9) called Interpolation in Figure 9:

$$A_{Sub} = \frac{3.79}{1+\exp(10.72*A_{Obj}-4.55)} + 1.13 \quad \text{with a correlation coefficient R=0.92} \qquad (9)$$

This interpolation allows to compare the subjective and the objective assessments relatively to the different filters. The Figure 10 illustrates for each filter, the subjective and the objective assessments. First, we note that the largest the filter is, the worst the quality is for both objective and subjective assessments. Then, by introducing the 95 percent confidence interval computed with the subjective results, we observe that the no-reference blur metric fits into this interval except for the nine pixels averaging filter that falls slightly outside.
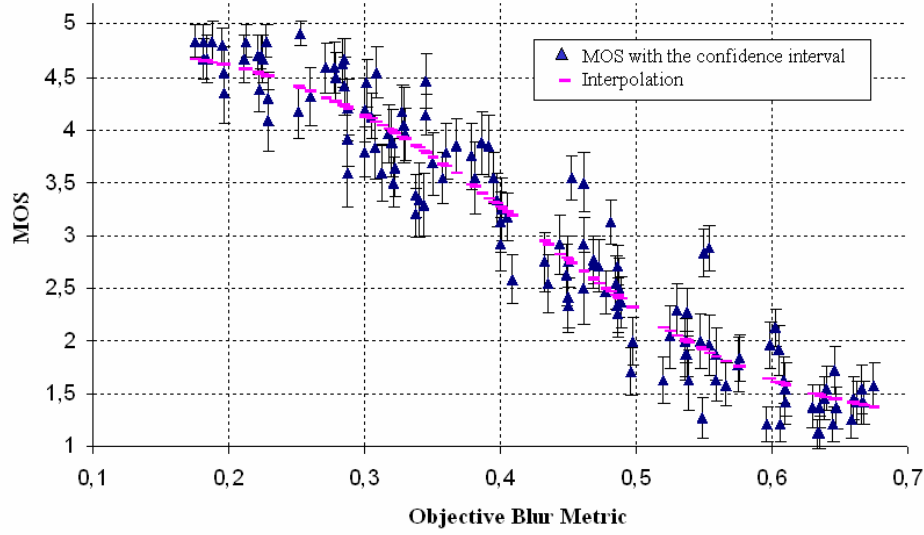
Figure 9: The no-reference perceptual blur metric versus the Mean Opinion Score (with the confidence interval)
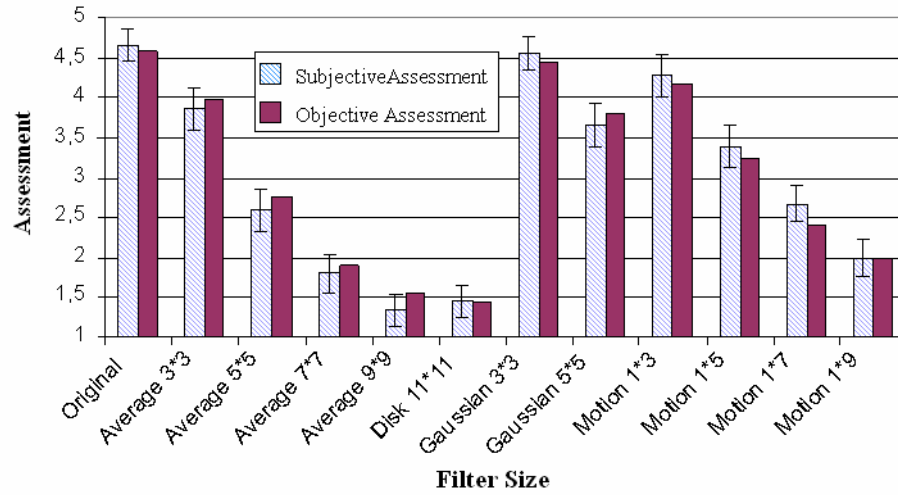

Figure 10: Influence of the filter sizes on the subjective assessment (with the confidence interval) and correlation with the objective metric interpolation

Note: In order to be a low-cost metric, we choose to compute the blur annoyance only in two directions (horizontal and vertical) but our principle is adapted to each directions. This choice depends not only to the wished precision but also to the cost implementation. In our case, we consider that the two directions give a result sufficiently accurate and well correlated with the human perception.

To conclude this subsection, we summarize in Figure 11 the several steps used to allow the validation of our no-reference blur quality metric as a no-reference perceptual blur quality metric.
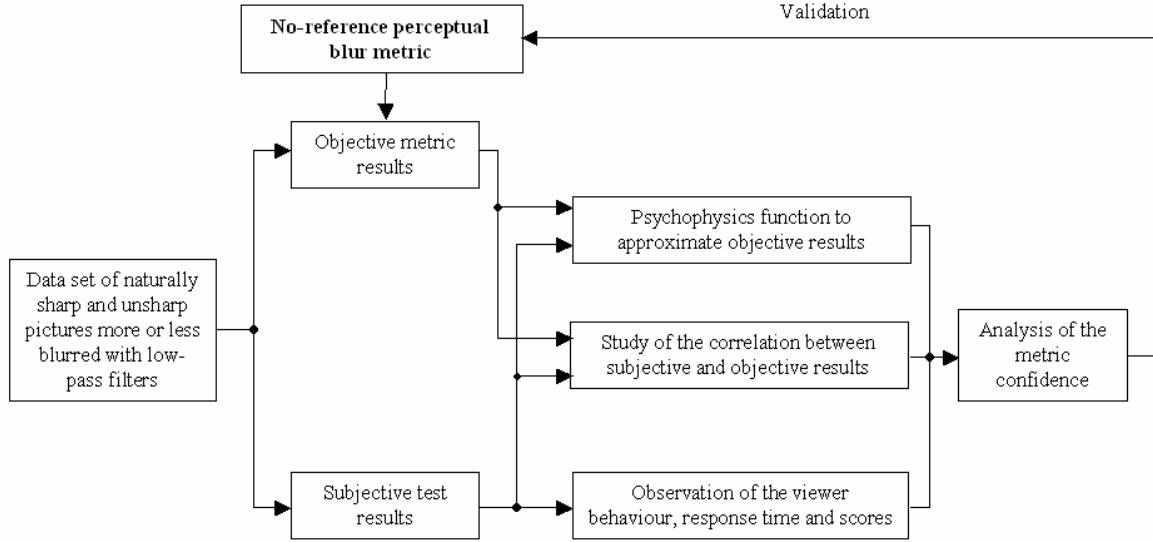
Figure 11: Flowchart of the no-reference blur quality metric validation step.

5. APPLICATIONS

This low-cost and robust no-reference perceptual blur metric can be used in different applications:

- A simple quality metric for pictures or video frames.

- A means to compare the quality of restoration methods or scaling methods. For example, this perceptual blur metric is well correlated with a subjective test aiming at estimate the quality of several scaling algorithms.

- A means to fine-tune artifacts correction algorithms which have a tendency to smooth details. For example, we use this algorithm to control the emergence of blur of a deblocking algorithm.

- An other application consists in the sharpness improvement by using an inverse method to evaluate the best appropriate sharpness filter from the blur coefficient

6. CONCLUSION

Based on the discrimination between different levels of blur perceptible on the same image, we propose a new method to estimate the blur annoyance. We justify each step of our method with a general observation on the blur perception and we develop a no-reference perceptual blur metric, ranging from 0 to 1 which are respectively the best and the worst quality in term of blur perception. To evaluate its performance, we compared this metric with subjective tests done following the ITU recommendations as good as possible. We validated this metric for static images by showing the high correlation between this no-reference blur metric and the human estimation.

This new measure can now be used as a blur quality metric but also as a means to improve a picture sharpness or to fine-tune artifact corrections avoiding the emergence of blur. About the subjective tests, it is interesting to notice that there was only one default introduced in the tested pictures. If one more default was introduced, the subjective measures could be less well aligned to the objective ones. Further experiments are currently in progress to assess several artifacts at the same time and to understand how to combine objective metrics of different artifacts to produce only one final quality factor.

ACKNOWLEDGMENTS

REFERENCES

1. Boult B.E., Chiang M.C. Local Blur Estimation and Super-Resolution. *In Proc. IEEE, Computer Society Conference on Computer Vision and Pattern Recognition*, pages 821–826, June 1997.
2. X. Marichal, Wei-Ying Ma, H.J. Zhang. Blur Determination in the Compressed Domain using DCT information. *In Proc. IEEE, International Conference on Image Processing*, 2:386–390, October 1999.
3. P. Marziliano, F. Dufaux, S. Winkler T. Ebrahimi. A No-Reference Perceptual Blur Metric. *In Proc. IEEE, International Conference on Image Processing*, 3:57–60, June 2002.
4. J. Caviedes, S. Gurbuz. No-reference sharpness metric based on local edge kurtosis. *In Proc. IEEE, International Conference on Image Processing*, 3:53–56, June 2002.
5. E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, F. Moschetti. A no-reference quality metric for measuring image blur. *In Proc. Seventh International Symposium on Signal Processing and Its Applications*, 1:469-472, July 2003.
6. H. Hu, G. de Haan, Low Cost Robust Blur Estimator. *In Proc. IEEE, International Conference on Image Processing,* October 2006.
7. ITU-R Recommandation BT.500-11. Methodology for the subjective assessment of the quality of television pictures. *In International Telecommunication Union*, 2002.