Blur-Invariant Similarity Measurement of Images

Matěj Lébl, Filip Šroubek, and Jan Flusser

Abstract—This article is a comment on the recent TPAMI paper (Gopalan *et al.*, 2012) that introduced a blur-invariant distance measure between two images. We point out two mistakes of the theory presented in (Gopalan *et al.*, 2012) and propose a correction. We also compare the original and corrected methods experimentally.

Index Terms—Blurred image classification, invariant distance, subspace projections, constrained minimization

1 Introduction

Blur is a common degradation factor of real digital images, which decreases significantly the recognition performance of image classification systems if not handled properly.

Capturing an ideal scene f by an imaging device with a point-spread function (PSF) h, the observed image g can be approximately modeled as a convolution

$$g = f * h, \tag{1}$$

(although in reality the degradation is usually much more complicated, Eq. (1) is a commonly accepted simplification). Estimating f is an ill-posed and time consuming problem that requires additional assumptions imposed on the PSF. This is why most recognition methods avoid image restoration and rather employ a blur-invariant image representation [1] or measure the similarity between images by means of a proper *blur-invariant distance*. Such a distance d should fulfill the constraint $d(f_1, f_2) = d(f_1 * h, f_2)$ for any admissible h. Since the latter approach is simpler from a theoretical point of view and is applicable to a broad family of PSFs, several authors have developed this idea recently [3], [4], [5], [6].

In this paper, we illustrate and correct two weaknesses of [3]. The first one is that [3] does not constrain the blurs to be physically realistic (non-negative and brightness-preserving). Although this limitation was mentioned in the text, it was not implemented in the algorithm. This weakness was later independently corrected in [5] by introducing additional constraints into the optimization. The second problem is more serious and has not been discussed in any paper. The method of [3] is actually not blur invariant, because its key theorem does not hold. In this paper, we expose where the mistake came from and present the correct solution.

2 Correcting Gopalan's Method

Gopalan *et al.* [3] derived a blur-invariant distance measure without assuming knowledge of the blur shape but they imposed a limitation on the blur support size. The authors showed that all blurred versions of the given image create a linear subspace, which can be understood as a point on a Grassmann manifold. The blur-invariant distance between two images is then defined as the Riemannian distance between two points on the manifold. At the same time, this can be equivalently understood as an angle between two subspaces. However, the method suffers from two major drawbacks. First, the quantity is claimed to be blur-invariant, but in reality it is not so. Second, the absence of any constraints imposed on the blur (except the support size) admits physically non-realistic blurs. To understand these drawbacks better, let us recall the Gopalan's method in more detail.

The central proposition in [3] is that the subspaces, generated by two images that differ from each other only by blur, are identical. The invariant property is constructed for a set of arbitrary blur kernels $\mathcal{H}$ with fixed support $m \times n = D$, i.e.,

$$\mathcal{H} = \{h \in (\mathbb{R})^{m\times n}\}. \tag{2}$$

Any blur $h \in \mathcal{H}$ can be written as a linear combination of orthonormal basis functions. For simplicity, we will assume the standard basis

$$e_{ij}(x, y) = \begin{cases} 1, & \text{if } x = i,\ y = j, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, n$. The convolution of an image f with e_{ij} generates a set of images of size $k \times l = N$, which we denote as

$$\mathcal{S}_f = \{f * e_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n\}. \tag{4}$$

The span of $\mathcal{S}_f$ contains results of the convolution of f with an arbitrary kernel of size up to $m \times n$, i.e., $\mathrm{span}(\mathcal{S}_f) = \{f * h | h \in \mathcal{H}\}$.

Gopalan's proposition states that $\mathrm{span}(\mathcal{S}_f)$ is a blur invariant of f to $\mathcal{H}$, since

$$\mathrm{span}(\mathcal{S}_f) = \mathrm{span}(\mathcal{S}_g), \text{ for } g = f * h,\ \forall h \in \mathcal{H}. \tag{5}$$

However, this is not true. The image g is blurred by h and in $\mathrm{span}(\mathcal{S}_g)$ it is assumed to be blurred again by another blur of size D. The final blur is not anymore in $\mathcal{H}$. We will now discuss the Gopalan's proof in more detail and state a corrected proposition.

Vectorizing elements of $\mathcal{S}_f$ and concatenating them, we construct a matrix $\mathbf{S}_f \in \mathbb{R}^{N\times D}$ of the form

$$\mathbf{S}_f = [\mathbf{F}\mathbf{e}_1, \ldots, \mathbf{F}\mathbf{e}_D] = \mathbf{F}\mathbf{E}, \tag{6}$$

where the matrix $\mathbf{F} \in \mathbb{R}^{N\times D}$ performs convolution with f, $\mathbf{F}\mathbf{e}_d$ is column vectorized $f * e_{ij}$ for $d = i + (j-1)m$, and $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_D]$ are column vectorized basis functions. For the chosen standard basis, the columns of $\mathbf{S}_f$ are spatially shifted versions of f cropped to size N. By definition, $\mathbf{E}$ is full rank and thus $\mathrm{span}(\mathbf{S}_f) = \mathrm{span}(\mathbf{F})$.

The proposition proof starts with considering an unknown blur kernel $b \in \mathcal{H}$ that produces g from a clean image f. From (4) follows that $\mathcal{S}_g = \{g * e_{ij}\} = \{f * b * e_{ij}\}$. The proof concludes that the vectorized form of $\mathcal{S}_g$ can be written as $\mathbf{S}_g = \mathbf{G}\mathbf{E} = \mathbf{F}\mathbf{B}\mathbf{E}$, where $\mathbf{B}$ is a matrix of size $D \times D$ performing convolution with b, and if $\mathbf{B}$ is full rank then $\mathrm{span}(\mathbf{S}_g) = \mathrm{span}(\mathbf{F}) = \mathrm{span}(\mathbf{S}_f)$.

Matrix $\mathbf{B}$ is of the Block-Toeplitz-Toeplitz-Block (BTTB) form and such matrices are indeed generally full rank. Gopalan's oversight is that $\mathbf{F}\mathbf{B} \neq \mathbf{G}$, where $\mathbf{G}$ similarly to $\mathbf{F}$ is of size $N \times D$ and performs convolution with g. Columns of convolution matrices $\mathbf{F}$ and $\mathbf{G}$ are shifted versions of f and g, respectively. For the product $\mathbf{F}\mathbf{B}$ to generate shifted versions of g, columns of $\mathbf{B}$ must contain shifted versions of b. However if the size of b is larger than one then $\mathbf{B}$ is rectangular and $\mathbf{F}$ in the product must have more columns than the original $\mathbf{F}$.

The equality of $\mathrm{span}(\mathbf{F})$ and $\mathrm{span}(\mathbf{G})$ is not true and instead the following proposition holds.

• *The authors are with the Czech Academy of Sciences, Institute of Information Theory and Automation, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic. E-mail: {lebl, sroubekf, flusser}@utia.cas.cz.*

Manuscript received 17 Apr. 2020; revised 27 Sept. 2020; accepted 2 Nov. 2020. Date of publication 6 Nov. 2020; date of current version 1 July 2021.
(Corresponding author: Jan Flusser.)
Recommended for acceptance by P. Favaro.
Digital Object Identifier no. 10.1109/TPAMI.2020.3036630

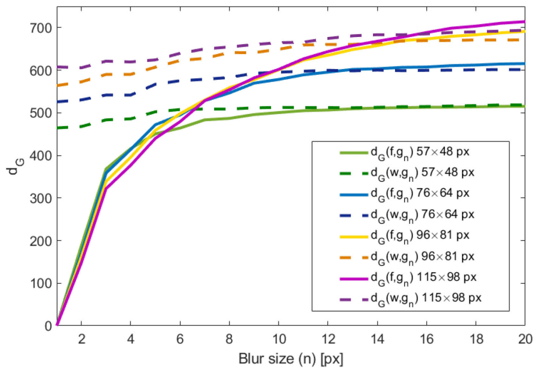

Fig. 1. The Gopalan's distance $d_G(f, g_k)$ (solid line) and $d_G(w, g_k)$ (dash line) as a function of k. According to [3], all solid lines should be constant zero.

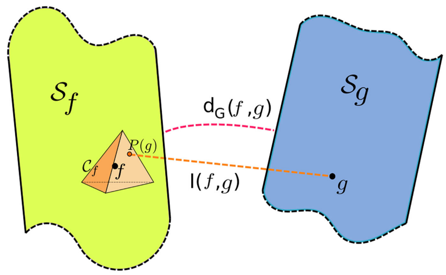

Fig. 2. Illustration of two between-image distance measures: Gopalan's "subspace to subspace" $d_G(f,g)$ and Vageeswaran's and Lébl's "image to a convex set" $I(f,g)$. Only the latter one is truly invariant to blur.

Proposition 1. *Let the size of the blur kernel b in $g = f * b$ be $D' = m' \times n'$, which is smaller that the original blur size $D = m \times n$, $m' \leq m$ and $n' \leq n$, then $\dim(\text{span}(\mathbf{F}) \cap \text{span}(\mathbf{G})) \geq D - D' + 1$.*

Proof. The convolution matrix $\mathbf{F}$ in (6) is of the form $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_D] \in \mathbb{R}^{N \times D}$, where $\mathbf{f}_i$ are vectorized shifted versions of f, and similarly $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_D]$ for g. We start by constructing a matrix $\bar{\mathbf{F}} = [\mathbf{f}_1, \ldots, \mathbf{f}_{\bar{D}}] \in \mathbb{R}^{N \times \bar{D}}$ that is similar to $\mathbf{F}$ but performs convolution of f with a blur of larger size $\bar{D} = (m + m' - 1) \times (n + n' - 1)$. Since $g = f * b$, every column of $\mathbf{G}$ is a linear combination of D' columns of $\bar{\mathbf{F}}$, i.e., $\mathbf{g}_i = [\mathbf{f}_i, \mathbf{f}_{i+1}, \ldots, \mathbf{f}_{i+D'-1}]\mathbf{b}$. Then $D' - 1$ vectors in $\mathbf{G}$ contain f_i's, which are not in $\mathbf{F}$ and therefore they may lie outside $\text{span}(\mathbf{F})$, and from this follows the lower bound of the intersection dimension. The equality holds if and only if $\bar{\mathbf{F}}$ has full column rank and all D' elements of b are nonzero. □

The Riemannian metric $\text{d}_G(f, g)$ used in the Gopalan's formulation measures the angle between the subspaces $\text{span}(\mathbf{S}_f)$ and $\text{span}(\mathbf{S}_g)$. If $\text{span}(\mathbf{F}) = \text{span}(\mathbf{G})$ then the angle is zero and $\text{d}_G(f, g) = 0$. The new proposition states that the subspaces are not equivalent in general, they only have a non-zero intersection. The Riemannian metric increases with the decreasing intersection dimension, i.e., d_G increases for g blurred with larger b.

We validated this result experimentally. We took two different images f and w (we used the YaleB dataset of face images [2]) and blurred f successively with a blur of a support D' ranging from 1×1 (no blur) to 20×20 (maximum blur, $D' = D$). In this way, we obtained blurred images $g_1, \ldots, g_{20}$. Then we measured the distance $\text{d}_G(f, g_k)$ which, according to the original Gopalan's proposition, should be equal to zero for each k. That is, however, not the case, as can be seen in Fig. 1. The solid line shows $\text{d}_G(f, g_k)$ as a function of k (four graphs, each for different choice of f). The distance is zero only for $k = 1$ (no blur) and then grows rapidly even for small blurs. We also evaluated the distance $\text{d}_G(w, g_k)$ for a comparison. It is visualized by the dash line (the solid and the dash lines of the same color belong to the same sequence of g_k's). This distance curve is more or less constant, as one may expect. For medium and large blurs, $\text{d}_G(f, g_k)$ gets close or even overpasses the distance $\text{d}_G(w, g_k)$, which leads to a misclassification in recognition experiments, as we demonstrate in Section 4.

3 Projection Method

Apart from the non-invariance of $\text{span}(\mathcal{S}_f)$, another conceptual problem in the Gopalan's approach is the omission of blur non-negativity and preservation of image energy. Gopalan *et al.* in [3] agree that physically plausible blur kernels have non-negative coefficients summing to one. The set of all plausible blurred images of f thus spans only a part of $\text{span}(\mathbf{S}_f)$. The Riemannian metric is defined only for subspaces and ignores the blur plausibility constraints. For this reason, the constraints are not enforced in [3].

The Gopalan's method was improved by Vageeswaran *et al.* [5], who imposed the blur plausibility constraints and defined the set of admissible blurs as

$$\mathcal{H} = \left\{ h \in (\mathbb{R})^{m \times n} \,\middle|\, h(x,y) \geq 0\,, \sum_{x,y} h(x,y) = 1 \right\}, \tag{7}$$

where $D = m \times n$ is again the maximal assumed blur size.

Under these constraints, blur-equivalent images form a convex set in the image space. The blur-invariant distance between the query image and the database template is defined as the distance between the point representing the query image, and its projection onto the convex set containing all blurred versions of the template

$$\mathcal{C}_f = \left\{ f * h \,\middle|\, h \in \mathcal{H} \right\}. \tag{8}$$

Most recently, essentially the same idea was independently proposed by Lébl *et al.* [4] who also presented an efficient algorithm for distance calculation by quadratic programming.

By definition, $g \in \mathcal{C}_f \Leftrightarrow g = h * f$ for some $h \in \mathcal{H}$. The blur invariant matching problem can be therefore seen as a task to find for a query image g the closest set $\mathcal{C}_f$.

Let us denote the projection of g onto $\mathcal{C}_f$ as $P_f(g)$. Then

$$I(f, g) = \|P_f(g) - g\|, \tag{9}$$

is a blur-invariant distance with respect to $\mathcal{H}$ defined in (7). The projection $P_f(g)$ is a point in $\mathcal{C}_f$ with the shortest euclidean distance from g. The blur-invariant distance can be found by minimization

$$I(f, g)^2 = \min_h \|f * h - g\|^2\,, \text{ s.t. } h \in \mathcal{H}\,, \tag{10}$$

which is a simple convex problem. If $\bar{h}$ denotes the minimizer of (10), then the projection to $\mathcal{C}_f$ is implemented as $P_f(g) = f * \bar{h}$. Let us remind the readers that $\bar{h}$ is found numerically without assuming any parametric shape or blur type.

Fig. 2 visualizes in a simplified way the difference between the two above mentioned distance measures. Note that $I(f, g)$ is not symmetric because the problem formulation itself is not symmetric – we assume that we know which image is clear (the one from the database) and which is blurred (the acquired one).

4 Experiments

We show that the Gopalan's distance (G) [3] leads to misclassifications if used for recognition of blurred images and the projection method (P) [4], [5][1] performs well since it is a true blur invariant.

1. These methods are equivalent in the part relevant to this study, both rely on minimization of Eq. (10). The code from [4] contains some speed-up tricks and is publicly available.

TABLE 1
Face Image Recognition

Blur size	6%	7%	9%	11%	13%	15%	17%
G rate [%]	100	98	94	85	79	78	52
P rate [%]	100	100	100	100	100	100	100

Success rate [%] of the Gopalan's method (G) and the projection algorithm (P) for varying motion blur. Blur size is given in % of the image size.

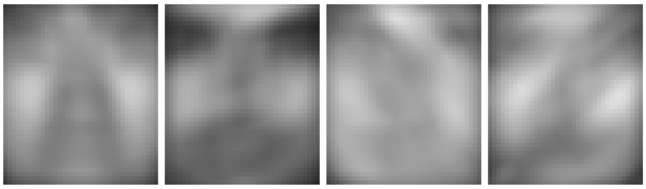

Fig. 3. Examples of faces recognized correctly by the projection method but misclassified by the Gopalan's method.

Fig. 4. Examples of original and blurred ImageNet pictures recognized by both methods (left), by the projection method only (middle) and an image not recognized by any method (right).

TABLE 2
Animal Image Recognition

Blur size	4%	5%	6%	7.5%	9%	12%
G rate, Gauss. blur [%]	100	100	88	59	42	21
P rate, Gauss. blur [%]	100	100	100	99	98	95
G rate, circular blur [%]	100	96	84	44	24	8
P rate, circular blur [%]	100	100	100	98	96	92

Success rate [%] of the Gopalan's method (G) and the projection algorithm (P) for varying Gaussian and circular out-of-focus blurs. Each success rate in the table was calculated from recognition of 300 query images.

TABLE 3
Time in Seconds Needed for a Single Image-to-Image Comparison for Gopalan's Method (G) and the Projection Algorithm (P) as a Function of the Image Size

Image size ($N \times N$ matrix)	22	32	37	42	47
G time [s]	0.07	0.22	0.92	2.3	5.01
P time [s]	0.001	0.004	0.005	0.002	0.01
Image size ($N \times N$ matrix)	57	62	67	72	82
G time [s]	22.04	33.2	55.6	82.1	199.6
P time [s]	0.005	0.05	0.003	0.02	0.03
Image size ($N \times N$ matrix)	87	92	97	107	117
G time [s]	294.9	368.6	504.7	933.9	1506
P time [s]	0.02	0.03	0.02	0.03	0.09

In the first experiment, similarly to [3], we used 38 face images of distinct persons from the YaleB dataset [2]. We blurred each of them twice by a simulated motion blur and recognized them against the originals. We repeated this seven times for blur size ranging from 6% of the image size to 17%, so we classified 532 blurred faces altogether by both methods. The recognition rates are summarized in Table 1. The projection method performs excellently on blurred images, there were no misclassifications at all. On the contrary, the performance of the Gopalan's method drops down as the blur increases, which is due to the effect analyzed in Section 2; see Fig. 3 for examples of blurred faces recognized correctly by P but misclassified by G.

The second experiment was performed on photographs downloaded from ImageNet dataset. We selected 200 images across various animal categories (dogs, cats, deer, cattle) and normalized them to the same graylevel mean and variance. Then we randomly selected 100 of them, blurred them and try to recognize them. This process was repeated for two blur types (Gaussian and circular out-of-focus blur) of the size ranging from 4% to 12% of the image size. Since we also introduced additive noise, each experiment was run three times. The recognition rates can be found in Table 2. Similarly to the previous experiment, we witness a significantly better performance of the projection method. Unlike the previous experiment, the projection method also yielded some misclassifications when the query image was heavily blurred (see Fig. 4) because there were similar "classes" in the dataset.

To conclude the experimental part, we compared both methods in terms of time complexity; see Table 3. The Gopalan's algorithm contains several time-consuming steps and it is thus much slower than the projection method, which is implemented using quadratic programming techniques.

Acknowledgments

This work was supported by the Czech Science Foundation under Grant No. GA18-07247S and by the Praemium Academiae.

References

[1] J. Flusser, T. Suk, J. Boldyš, and B. Zitová, "Projection operators and moment invariants to image blurring," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 4, pp. 786–802, Apr. 2015.

[2] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intel.* vol. 23, no. 6, pp. 643–660, Jun. 2001.

[3] R. Gopalan, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," *IEEE Trans. Pattern Anal. Mach. Intel.* vol. 34, no. 6, pp. 1220–1226, Jun. 2012.

[4] M. Lébl, F. Šroubek, J. Kautský, and J. Flusser, "Blur invariant template matching using projection onto convex sets," in *Proc. 18th Int. Conf. Comput. Anal. Images Patterns*, 2019, pp. 351–362.

[5] P. Vageeswaran, K. Mitra, and R. Chellappa, "Blur and illumination robust face recognition via set-theoretic characterization," *IEEE Trans. Image Process.* vol. 22, no. 4, pp. 1362–1372, Apr. 2013.

[6] Z. Zhang, E. Klassen, and A. Srivastava, "Gaussian blurring-invariant comparison of signals and images," *IEEE Trans. Image Process.* vol. 22, no. 8, pp. 3145–3157, Aug. 2013.