Template Matching with Noisy Patches: A Contrast-Invariant GLR Test

Charles-Alban Deledalle, Loïc Denis, Florence Tupin

TEMPLATE MATCHING WITH NOISY PATCHES: A CONTRAST-INVARIANT GLR TEST

Charles-Alban Deledalle

IMB
CNRS-Univ. Bordeaux 1
France

Loïc Denis

Laboratoire Hubert Curien
CNRS-Univ. Saint-Etienne
France

Florence Tupin

Télécom ParisTech
CNRS-LTCI
France

ABSTRACT

Matching patches from a noisy image to atoms in a dictionary of patches is a key ingredient to many techniques in image processing and computer vision. By representing with a single atom all patches that are identical up to a radiometric transformation, dictionary size can be kept small, thereby retaining good computational efficiency. Identification of the atom in best match with a given noisy patch then requires a contrast-invariant criterion. In the light of detection theory, we propose a new criterion that ensures contrast invariance and robustness to noise. We discuss its theoretical grounding and assess its performance under Gaussian, gamma and Poisson noises.

Index Terms— Template matching, Likelihood ratio test, Detection theory, Image restoration

1. INTRODUCTION

In this paper, we address the problem of template matching of patches under various noise conditions. More precisely, when provided a collection of noise-free templates (the dictionary), we focus on finding for a given noisy patch the best matching element in the dictionary. Template matching is at the heart of many recent image processing and computer vision techniques, for instance, for denoising [1] or classification with a labeled dictionary [2]. We focus in the following on how to perform template matching when the noise departs from the Gaussian distribution. Inspired by our previous work about the comparison of noisy patches [3], we extend here the proposed methodology to the problem of template matching.

By $\boldsymbol{x}$, we denote a patch of an image, i.e., a collection of N noisy pixel values. By $\boldsymbol{a} \in \mathcal{D}$, we denote a template taken from a dictionary $\mathcal{D}$ ($\boldsymbol{a}$ also has N pixels). We do not specify here a shape but consider that the values are ordered so that when a patch $\boldsymbol{x}$ is compared to a template $\boldsymbol{a}$, values with identical index are in spatial correspondence. For best efficiency, dictionaries should be as small as possible while being representative of images. To limit the size of dictionaries, a common idea is to let atoms represent a class of patches that are identical up to a radiometric transformation. Hence, a template should essentially encode the geometrical patterns of

a patch rather than its radiometry. Of course, to exploit such a dictionary, the template matching criterion must be invariant to the radiometric changes considered while being robust to the noise statistic.

We assume that the noise can be modeled by a (known) distribution so that a noisy patch $\boldsymbol{x}$ is a realization of an N-dimensional random variable $\boldsymbol{X}$ modeled by a probability density or mass function $p(\cdot|\boldsymbol{\theta})$. The vector of parameters $\boldsymbol{\theta}$ is referred in the following as the noise-free patch. For example, a patch $\boldsymbol{x}$ damaged by additive white Gaussian noise with standard deviation σ can be modeled by:

$$\boldsymbol{x} = \boldsymbol{\theta} + \sigma \boldsymbol{n} \qquad (1)$$

where $\boldsymbol{\theta}$ is the noise-free patch and $\boldsymbol{n}$ is the realization of a zero-mean normalized Gaussian random vector with independent elements. It is straightforward to see that $\boldsymbol{X}|\boldsymbol{\theta}$ follows a Gaussian distribution with mean $\boldsymbol{\theta}$ and standard deviation σ. While such decompositions exist for some specific distributions (e.g., gamma distribution involves a multiplicative decomposition), in most cases no decomposition of $\boldsymbol{x}$ in terms of $\boldsymbol{\theta}$ and an independent noise component may be found (e.g., under Poisson noise). In general, when noise departs from additive Gaussian noise, the link between $\boldsymbol{X}$ and $\boldsymbol{\theta}$ is described by the probability density or mass function $p(\boldsymbol{x}|\boldsymbol{\theta})$.

2. PROBLEM DEFINITION

A template matching criterion c defines a mapping from a pair formed by a noisy patch and a template $(\boldsymbol{x}, \boldsymbol{a})$ to a real value. The larger the value of $c(\boldsymbol{x}, \boldsymbol{a})$, the more relevant the match between $\boldsymbol{x}$ and the template $\boldsymbol{a}$. We consider that a matching criterion c is invariant with respect to the family of transformations $\mathcal{T}_{\boldsymbol{\rho}}$ parametrized by vector $\boldsymbol{\rho}$, if

$$\forall \boldsymbol{X}, \boldsymbol{a}, \boldsymbol{\rho}, \quad c(\boldsymbol{X}, \mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a})) = c(\boldsymbol{X}, \boldsymbol{a}) \,.$$

A typical example is to consider invariance up to an affine change of contrast: $\mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a}) = \mathcal{T}_{\alpha,\beta}(\boldsymbol{a}) = \alpha \boldsymbol{a} + \beta \mathbb{1}$, where $\mathbb{1}_k = 1$ for all $1 \le k \le N$. In the light of detection theory, we consider that a noisy patch $\boldsymbol{x}$ and a template $\boldsymbol{a}$ are in match (up to a transformation $\mathcal{T}_{\boldsymbol{\rho}}$) when $\boldsymbol{x}$ is a realization of a random variable $\boldsymbol{X}$ following a distribution $p(.|\boldsymbol{\theta})$ for which

there exists a vector of parameters $\boldsymbol{\rho}$ such that $\boldsymbol{\theta} = \mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a})$. The template matching problem can then be rephrased as the following hypothesis test (a parameter test):

$$\begin{aligned} \mathcal{H}_0 &: \exists \boldsymbol{\rho} & \boldsymbol{\theta} = \mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a}) & \quad \text{(null hypothesis)}, \\ \mathcal{H}_1 &: \forall \boldsymbol{\rho} & \boldsymbol{\theta} \neq \mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a}) & \quad \text{(alternative hypothesis)}. \end{aligned}$$

For a given template matching criterion c, the *probability of false alarm* (to decide $\mathcal{H}_1$ under $\mathcal{H}_0$) and the *probability of detection* (to decide $\mathcal{H}_1$ under $\mathcal{H}_1$) are defined as:

$$P_{FA} = \mathbb{P}(c(\boldsymbol{X}, \boldsymbol{a}) < \tau | \boldsymbol{\rho}, \mathcal{H}_0), \qquad (2)$$
$$P_D = \mathbb{P}(c(\boldsymbol{X}, \boldsymbol{a}) < \tau | \boldsymbol{\theta}, \mathcal{H}_1). \qquad (3)$$

Note that the inequality symbols are reversed compared to usual definitions since we consider detection of mismatch based on the matching measure c.

According to Neyman-Pearson theorem, the optimal criterion, i.e., the criterion which maximizes P_D for any given P_{FA}, is the likelihood ratio (LR) test:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{a}) = \frac{p(\boldsymbol{x} | \boldsymbol{\theta} = \mathcal{T}_{\boldsymbol{\rho}}(\boldsymbol{a}), \mathcal{H}_0)}{p(\boldsymbol{x} | \boldsymbol{\theta}, \mathcal{H}_1)}. \qquad (4)$$

The application of the likelihood ratio test requires the knowledge of $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ (the parameters of the transformation and the noise-free patch) which, of course, are unavailable. Our problem is thus a *composite hypothesis problem*. A criterion maximizing P_D for all P_{FA} and all values of the unknown parameters is said *uniformly most powerful* (UMP). Kendall and Stuart (1979) showed that no UMP detector exists in general for our *composite hypothesis problem* [4], so that any criteria can be defeated by another one at a specific P_{FA}. The research of a universal template matching criterion is then futile. We address here the question of how different criteria behave on patches extracted from natural images.

3. CONTRAST-INVARIANT TEMPLATE MATCHING

In this section we consider radiometric changes $\mathcal{T}_{\alpha,\beta}$ defined by two parameters: α and β. We present different candidate criteria for contrast-invariant template matching and discuss their robustness to the noise statistics.

Normalized correlation: The most usual way to measure similarity up to an affine change of contrast of the form $\mathcal{T}_{\alpha,\beta}(\boldsymbol{x}) = \alpha \boldsymbol{x} + \beta \mathbb{1}$ between two (non-constant) vectors $\boldsymbol{x}$ and $\boldsymbol{a}$ is to consider their normalized correlation:

$$\mathcal{C}(\boldsymbol{x}, \boldsymbol{a}) = \left| \frac{\sum_{k=1}^{N} (x_k - \bar{x})(a_k - \bar{a})}{\sqrt{\sum_{k=1}^{N} (x_k - \bar{x})^2 \sum_{k=1}^{N} (a_k - \bar{a})^2}} \right|. \qquad (5)$$

where $\bar{x} = \frac{1}{N} \sum_k x_k$ and $\bar{a} = \frac{1}{N} \sum_k a_k$. Indeed, it is straightforward to show that the correlation provides the desired contrast invariance property. Regarding noise corruptions, it is not straightforward whether the correlation is a robust template matching criterion. We will show that, under

the assumption of Gaussian noise, for a fixed observation $\boldsymbol{x}$, the vector $\boldsymbol{a} \in \mathcal{D}$ that maximizes the correlation also maximizes the likelihood up to an affine change of contrast.

Generalized Likelihood Ratio: Motivated by optimality guarantees of the LR test (4) and our previous work in [3], a template matching criterion can be defined from statistical detectors designed for *composite hypothesis problems*. The generalized LR (GLR) replaces the unknowns α, β and $\boldsymbol{\theta}$ in eq. (4) by their maximum likelihood estimates (MLE) under each hypothesis:

$$\begin{aligned} \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) &= \frac{\sup_{\alpha,\beta} p(\boldsymbol{x} | \boldsymbol{\theta} = \mathcal{T}_{\alpha,\beta}(\boldsymbol{a}), \mathcal{H}_0)}{\sup_{\boldsymbol{t}} p(\boldsymbol{x} | \boldsymbol{\theta} = \boldsymbol{t}, \mathcal{H}_1)} \\ &= \frac{p(\boldsymbol{x} | \boldsymbol{\theta} = \mathcal{T}_{\hat{\alpha},\hat{\beta}}(\boldsymbol{a}))}{p(\boldsymbol{x} | \boldsymbol{\theta} = \hat{\boldsymbol{t}})} \end{aligned} \qquad (6)$$

where $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\boldsymbol{t}}$ are the MLE of the unknown α, β and $\boldsymbol{\theta}$. By construction, the GLR satisfies the contrast invariance property. Asymptotically to the SNR, GLR is optimal due to the efficiency of MLE. Its asymptotic distribution is known and so are the P_{FA} values associated to any given threshold τ: GLR is asymptotically a constant false alarm rate (CFAR) detector. The GLR test is also invariant upon changes of variable [5]: it does not depend on the representation of the noisy patch. While we noted that there are no UMP detectors for our composite hypothesis problem, GLR is *asymptotically* UMP among invariant tests [6]. Due to its dependency on MLE, the performance of GLR may fall in low SNR conditions, where the MLE is known to behave poorly.

Stabilization: A classical approach to extend the applicability of a matching criterion to non-Gaussian noises is to apply a transformation to the noisy patches. The transformation is chosen so that the transformed patches follow a (close to) Gaussian distribution with constant variance (hence their name: variance-stabilization transforms). This leads for instance to the homomorphic approach which maps multiplicative noise to additive noise with stationary variance. This is also the principle of Anscombe transform and its variants used for Poisson noise. Given an application s which stabilizes the variance for a specific noise distribution, stabilization-based criteria can be obtained using (5) or (6) on the output of s:

$$\mathcal{S}_{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{a}) = \mathcal{C}(s(\boldsymbol{x}), s(\boldsymbol{a})), \qquad (7)$$
$$\mathcal{S}_{\mathcal{G}}(\boldsymbol{x}, \boldsymbol{a}) = \mathcal{G}(s(\boldsymbol{x}), s(\boldsymbol{a})) \qquad (8)$$

where the likelihood function $p(s(\boldsymbol{x}) | s(\boldsymbol{\theta}))$ is assumed to be a Gaussian distribution centered on $s(\boldsymbol{\theta})$ with a covariance matrix $\sigma^2 \mathbf{I}$. As we will see, an advantage of this approach compared to the GLR criterion is that it is usually simpler to evaluate in closed-form, and then, leads to faster algorithms. An important limitation of this approach lies nevertheless in the existence of a stabilization function s. Beyond existence, the performance of this approach may fall if the transformed data distribution is far from the Gaussian distribution.

4. GLR IN DIFFERENT NOISE CONDITIONS

In this section, we provide closed-form expressions or iterative schemes to evaluate the GLR in the case of Gaussian noise, gamma noise and Poisson noise.

Proposition 1 (Gaussian noise)**.** *Consider that $\boldsymbol{X}$ follows a Gaussian distribution such that*

$$p(x_k|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_k - \theta_k)^2}{2\sigma^2}\right) ,$$

and consider the class of affine transformations $\mathcal{T}_{\alpha,\beta}(\boldsymbol{x}) = \alpha\boldsymbol{x} + \beta\mathbb{1}$. In this case, we have

$$-\log\mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = (1 - \mathcal{C}(\boldsymbol{x}, \boldsymbol{a})^2)\frac{\|\boldsymbol{x} - \bar{x}\mathbb{1}\|_2^2}{2\sigma^2} .$$

Proof. For the Gaussian law, the MLE of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{t}} = \boldsymbol{x}$ so that

$$-\log\mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = \frac{\|x - \hat{\alpha}\boldsymbol{a} - \hat{\beta}\mathbb{1}\|_2^2}{2\sigma^2}$$

and $\hat{\alpha}$ and $\hat{\beta}$ are the coefficients of the linear least squared regression, i.e.,

$$\hat{\alpha} = \frac{\sum_{k=1}^{N}(x_k - \bar{x})(a_k - \bar{a})}{\sum_{k=1}^{N}(a_k - \bar{a})^2} \quad \text{and} \quad \hat{\beta} = \bar{x} - \alpha\bar{a} ,$$

with $\bar{x}$ and $\bar{a}$ the empirical mean of $\boldsymbol{x}$ and $\boldsymbol{a}$. Injecting the expression of $\hat{\alpha}$ and $\hat{\beta}$ in the previous equation gives the proposed formula. $\square$

Remark that for a fixed observation $\boldsymbol{x}$ and any $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathcal{D}$, if $\mathcal{C}(\boldsymbol{x}, \boldsymbol{a}_1) < \mathcal{C}(\boldsymbol{x}, \boldsymbol{a}_2)$ then $\mathcal{G}(\boldsymbol{x}, \boldsymbol{a}_1) < \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}_2)$. In particular, we have

$$\operatorname*{argmax}_{\boldsymbol{a}\in\mathcal{D}} \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = \operatorname*{argmax}_{\boldsymbol{a}\in\mathcal{D}} \mathcal{C}(\boldsymbol{x}, \boldsymbol{a})$$
$$= \operatorname*{argmax}_{\boldsymbol{a}\in\mathcal{D}} \sup_{\alpha,\beta} p(\boldsymbol{x}|\boldsymbol{\theta} = \mathcal{T}_{\alpha,\beta}(\boldsymbol{a}), \mathcal{H}_0)$$

which is the MLE under the hypothesis $\mathcal{H}_0$. However, beyond equivalence of their maxima, the GLR is not equivalent to the correlation even in the case of Gaussian noise. They have different detection performance when the purpose is to take a decision by thresholding their answer. Compared to the correlation, GLR adapts its answer with respect to $\frac{\|\boldsymbol{x}-\bar{x}\mathbb{1}\|^2}{2\sigma^2}$ which, in some sense, measures the signal-to-noise-ratio (SNR) in $\boldsymbol{x}$. For a fixed threshold τ and $\boldsymbol{a}$, if the SNR of $\boldsymbol{x}$ is small enough, GLR will put the pair $(\boldsymbol{x}, \boldsymbol{a})$ in correspondence whatever their content. In fact, when the SNR is small enough, any template up to a radiometric transform can explain the observed realization. The correlation, which does not take into account the noise in its definition, does not adapt to the SNR of $\boldsymbol{x}$. Worse, the correlation tends to increase when the SNR of $\boldsymbol{x}$ decreases. We will see in Section 4 that such a behavior of GLR is of main importance for a template matching task.

Proposition 2 (Gamma noise)**.** *Consider that $\boldsymbol{X}$ follows a gamma distribution such that*

$$p(x_k|\theta_k) = \frac{L^L x_k^{L-1}}{\Gamma(L)\theta_k^L} \exp\left(-\frac{Lx_k}{\theta_k}\right)$$

and consider the class of log-affine transformations $\mathcal{T}_{\alpha,\beta}(\boldsymbol{x}) = \beta\boldsymbol{x}^\alpha$ where $(.)^\alpha$ is the element-wise power function. In this case, we have

$$-\log\mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = L\sum_{k=1}^{N}\log\left(\frac{\hat{\beta}a_k^{\hat{\alpha}}}{x_k}\right)$$

where $\hat{\alpha}$ and $\hat{\beta}$ can be obtained iteratively as

$$\hat{\alpha}_{i+1} = \hat{\alpha}_i - \frac{\sum_k(1 - r_{k,i})\log a_k}{\sum_k r_{k,i}(\log a_k)^2} \quad \text{and} \quad \hat{\beta}_{i+1} = \frac{1}{N}\sum_k \frac{x_k}{a_k^{\hat{\alpha}_i}}$$

with $r_{k,i} = x_k/(\hat{\beta}_i a_k^{\hat{\alpha}_i})$, whatever the initialization.

Proof. For the gamma law, the MLE of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{t}} = \boldsymbol{x}$ so that

$$-\log\mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = L\sum_{k=1}^{N}\left(\log\frac{\hat{\beta}a_k^{\hat{\alpha}}}{x_k} + \frac{x_k}{\hat{\beta}a_k^{\hat{\alpha}}} - 1\right) .$$

The function $\beta \mapsto \sum_k -\log p(x_k|\beta a_k^\alpha)$ has a unique minimum at $\frac{1}{N}\sum_k \frac{x_k}{a_k^\alpha}$. Moreover, the function $\alpha \mapsto \sum_k -\log p(x_k|\theta_k = \beta a_k^\alpha)$ is convex and twice differentiable, therefore the Newton method can be used to estimate $\hat{\alpha}$ whatever the initialization. Differentiating twice $\alpha \mapsto \sum_k -\log p(x_k|\theta_k = \beta a_k^\alpha)$ gives the proposed iterative scheme. Injecting the value of $\hat{\alpha}$ and $\hat{\beta}$ in the previous equation gives the proposed formula. $\square$

Unlike in the case of the Gaussian law, there is no closed-form formula of GLR in the case of the gamma law and one should rather compute it iteratively. Note that in practice only a few iterations are required if one initializes using the log-moment estimation, as suggested in [7], leading to the following initialization:

$$\hat{\alpha}_0 = \sqrt{\frac{\max(\sum_k(\log x_k - \overline{\log x})^2 - \psi(1, L), 0)}{\sum_k(\log a_k - \overline{\log a})^2}}$$
$$\hat{\beta}_0 = \exp\left(\overline{\log x} - \psi(L) + \log(L) - \alpha\overline{\log a}\right)$$

where $\overline{\log x} = \frac{1}{N}\sum_k \log x_k$ and $\overline{\log a} = \frac{1}{N}\sum_k \log a_k$.

Proposition 3 (Poisson noise)**.** *Consider that $\boldsymbol{X}$ follows a Poisson distribution so that*

$$p(x_k|\theta_k) = \frac{\theta_k^{x_k} e^{-\theta_k}}{x_k!}$$

and consider the class of log-affine transformations $\mathcal{T}_{\alpha,\beta}(\boldsymbol{x}) = \beta\boldsymbol{x}^\alpha$. In this case, we have
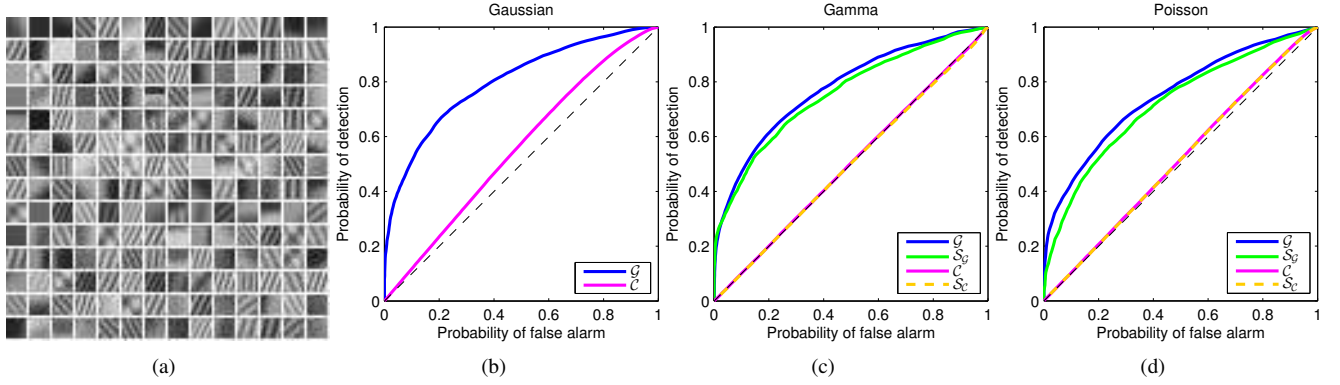
Fig. 1. (a) Patch dictionary. (b) ROC curve obtained under Gaussian noise, (c) ROC curve obtained under gamma noise and (d) ROC curve obtained under Poisson noise. In all experiments, the SNR over the whole dictionary is about -3dB.

$$-\log \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = \sum_{k=1}^{N} x_k \log \left(\frac{x_k}{\hat{\beta} a_k^{\hat{\alpha}}} \right)$$

where $\hat{\alpha}$ and $\hat{\beta}$ can be obtained iteratively as

$$\hat{\alpha}_{i+1} = \hat{\alpha}_i - \frac{\sum_k \left(\hat{\beta}_i a_k^{\hat{\alpha}_i} - x_k \right) \log a_k}{\sum_k \hat{\beta}_i a_k^{\hat{\alpha}_i} (\log a_k)^2} \quad and \quad \hat{\beta}_{i+1} = \frac{\sum_k x_k}{\sum_k a_k^{\hat{\alpha}_i}}$$

whatever the initialization.

Proof. For the Poisson law, the MLE of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{t}} = \boldsymbol{x}$ such that

$$-\log \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) = \sum_{k=1}^{N} \left(x_k \log \left(\frac{x_k}{\hat{\beta} a_k^{\hat{\alpha}}} \right) + \hat{\beta} a_k^{\hat{\alpha}} - x_k \right).$$

The function $\beta \mapsto \sum_k -\log p(x_k|\beta a_k^\alpha)$ has a unique minimum at $\frac{\sum_k x_k}{\sum_k a_k^\alpha}$. Moreover, the function $\alpha \mapsto \sum_k -\log p(x_k|\theta_k = \beta a_k^\alpha)$ is convex and twice differentiable, such that the Newton method can be used to estimate $\hat{\alpha}$ whatever the initialization. Differentiating twice $\alpha \mapsto \sum_k -\log p(x_k|\theta_k = \beta a_k^\alpha)$ gives the proposed iterative scheme. Injecting the value of $\hat{\alpha}$ and $\hat{\beta}$ in the previous equation gives the proposed formula. □

Again there is no closed-form formula of GLR, but in practice only a few iterations are required if one uses the $\hat{\alpha}$ and $\hat{\beta}$ that minimize the linear least square error between $\log \boldsymbol{x}$ and $\log \boldsymbol{a}$.

5. EVALUATION OF PERFORMANCE

5.1. Detection performance

We evaluate the relative performance of the correlation, GLR and the variance stabilization based matching criteria on a dictionary composed of 196 noise-free patches of size $N = 8 \times 8$. The noise-free patches have been obtained using the k-means

on patches extracted from the classical 512×512 *Barbara* image. Each noisy patch $\boldsymbol{x}$ is a noisy realization of the noise-free patches under Gaussian, gamma or Poisson noise with an overall SNR of about -3dB. Each template $\boldsymbol{a}$ is a randomly transformed atom of the dictionary up to an affine change of contrast for the experiments involving Gaussian noise, and up to a log-affine change of contrast under gamma or Poisson noises. All criteria are evaluated for all pairs $(\boldsymbol{x}, \boldsymbol{a})$. The process is repeated 20 times with independent noise realizations and radiometric transformations.

The performance of the matching criteria is given in term of their *receiver operating characteristic* (ROC) curve, i.e., the curve of P_D with respect to P_{FA}, where we have relaxed the hypothesis test as

$$\begin{aligned} \mathcal{H}_0 &: \exists \alpha, \beta & \boldsymbol{\theta} \approx \mathcal{T}_{\alpha,\beta}(\boldsymbol{a}) & \quad \text{(null hypothesis)}, \\ \mathcal{H}_1 &: \forall \alpha, \beta & \boldsymbol{\theta} \not\approx \mathcal{T}_{\alpha,\beta}(\boldsymbol{a}) & \quad \text{(alternative hypothesis)} \end{aligned}$$

and where $\boldsymbol{\theta} \approx \mathcal{T}_{\alpha,\beta}(\boldsymbol{a})$ reads as: on average, the noise-free patch $\mathcal{T}_{\alpha,\beta}(\boldsymbol{a})$ explains almost as well the realizations of $\boldsymbol{X}$ than the actual noise-free patch $\boldsymbol{\theta}$, and is measured by:

$$\mathcal{D}_{KL}(\boldsymbol{\theta} \parallel \mathcal{T}_{\alpha,\beta}(\boldsymbol{a})) \leq \nu,$$

where $\mathcal{D}_{KL}$ is the Kullback-Leibler divergence and ν is a small value (chosen here equal to 0.02). Results are given in Figure 1. Even with Gaussian noise or with variance stabilization, the correlation behaves poorly in noisy condition. The generalized likelihood ratio (GLR) is the most powerful criterion followed by the GLR with variance stabilization.

5.2. Application to dictionary-based denoising

We exemplify here the performance of GLR in a dictionary-based denoising task. The dictionary $\mathcal{D}$ is considered describing a generative model of the patches $\boldsymbol{x}$ of the noisy image as realizations of $\boldsymbol{X}$ following a distribution of parameter $\boldsymbol{\theta} = \mathcal{T}_{\alpha,\beta}(\boldsymbol{a})$ with $\boldsymbol{a} \in \mathcal{D}$. Under this model, we suggest estimating each patch of the image as:
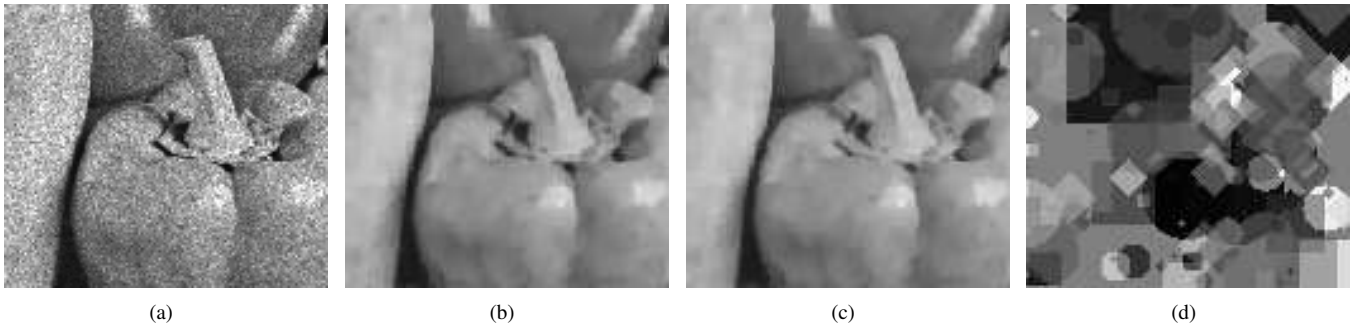
Fig. 2. (a) Noisy input image damaged by gamma noise (PSNR=21.14). (b) Denoised image using the GLR after variance stabilization (PSNR=27.42). (c) Denoised image using the GLR adapted to gamma noise (PSNR=27.53). (d) Image composed of the atoms of the dictionary.

$$\hat{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{1}{Z} \sum_{\boldsymbol{a} \in \mathcal{D}} \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}) \boldsymbol{a}^{\star} \quad \text{with} \quad Z = \sum_{\boldsymbol{a} \in \mathcal{D}} \mathcal{G}(\boldsymbol{x}, \boldsymbol{a}), \quad (9)$$

where $\boldsymbol{a}^{\star} = \mathcal{T}_{\hat{\alpha}, \hat{\beta}}(\boldsymbol{a})$ and $\hat{\alpha}$ and $\hat{\beta}$ are the MLE of α and β used in the calculation of $\mathcal{G}(\boldsymbol{x}, \boldsymbol{a})$. Equation (9) has a Bayesian interpretation as the posterior mean estimator:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{a} \in \mathcal{D}} p(\boldsymbol{a}^{\star}|\boldsymbol{x}) \boldsymbol{a}^{\star}}{\sum_{\boldsymbol{a} \in \mathcal{D}} p(\boldsymbol{a}^{\star}|\boldsymbol{x})}, \quad (10)$$

considering a priori that the frequencies of the atoms of $\mathcal{D}$ are uniform in the image. The posterior mean is known to minimize the Bayesian least square error $\mathbb{E}\left[\|\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2 \mid \boldsymbol{\theta}\right]$.

Figure 2 shows the denoising results obtained on a 128×128 image damaged by gamma noise (with $L = 10$) using (9) with the GLR adapted to gamma noise and with the GLR adapted to a Gaussian law after variance stabilization[1]. The dictionary $\mathcal{D}$ is chosen as the set of all atoms extracted from a 128×128 image (a.k.a., an epitome) built following the transparent dead leaves model of [9]. This model ensures the dictionary to be shift invariant [10, 11] while representing information of different scales. As in [10, 11], we manipulate epitomes in Fourier domain in order to evaluate eq. (9) efficiently. Eventually, Fig. 2 shows that using the GLR for the gamma law or for the Gaussian law after stabilizing the variance are both satisfactory visually and in term of PSNR.

6. CONCLUSION

Normalized correlation is widely used as a contrast-invariant criterion for template matching. We have shown that the GLR test provides a criterion that is more robust to noise. In the case of Gaussian noise, this criterion involves both a normalized correlation term and a term that evaluates the signal-to-noise ratio of the noisy data. Under non-Gaussian noise distributions, criteria derived from the GLR test are generally not known in closed form but require a few iterations to be evaluated. When variance stabilization technique can be employed,

---

[1]when using stabilization, a debiasing step is performed following [8].

our numerical experiments show that good performance is reached using Gaussian GLR after variance stabilization.

7. REFERENCES

[1] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[2] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 791–804, 2012.

[3] C-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? patch similarity beyond gaussian noise," *International journal of computer vision*, pp. 1–17, 2012.

[4] M. Kendall and A. Stuart, "The advanced theory of statistics. Vol. 2: Inference and relationship," 1979.

[5] S.M. Kay and J.R. Gabriel, "An invariance property of the generalized likelihood ratio test," *IEEE Signal Processing Letters*, vol. 10, no. 12, pp. 352–355, 2003.

[6] E.L. Lehmann, "Optimum invariant tests," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 881–884, 1959.

[7] J-M. Nicolas, "Introduction to second kind statistics: application of log-moments and log-cumulants to SAR image law analysis," *Traitement du Signal*, vol. 19, no. 3, pp. 139–168, 2002.

[8] H. Xie, LE Pierce, and FT Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196–2212, 2002.

[9] B. Galerne and Y. Gousseau, "The transparent dead leaves model," *Advances in Applied Probability*, vol. 44, no. 1, pp. 1–20, 2012.

[10] P. Jost, P. Vandergheynst, S. Lesage, and R. Gribonval, "Motif: an efficient algorithm for learning translation invariant dictionaries," in *Proceedings of ICASSP 2006*. IEEE, 2006.

[11] L. Benoît, J. Mairal, F. Bach, and J. Ponce, "Sparse image representation with epitomes," in *Proceedings of CVPR, 2011*. IEEE, 2011.