Bivariate Gamma Distributions for Image Registration and Change Detection

Florent Chatelain[†], Jean-Yves Tourneret[†], Jordi Inglada[•] and André Ferrari[*]

[†]IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

[•]CNES, 18 avenue E. Belin, BPI 1219, 31401 Toulouse cedex 9, France

[*]LUAN, Université de Nice-Sophia-Antipolis, 06108 Nice cedex 2, France

florent.chatelain@enseeiht.fr, jean-yves.tourneret@enseeiht.fr, jordi.inglada@cnes.fr, ferrari@unice.fr

Abstract— This paper evaluates the potential interest of using bivariate gamma distributions for image registration and change detection. The first part of the paper studies estimators for the parameters of bivariate gamma distributions based on the maximum likelihood principle and the method of moments. The performance of both methods are compared in terms of estimated mean square errors and theoretical asymptotic variances. The mutual information is a classical similarity measure which can be used for image registration or change detection. The second part of the paper studies some properties of the mutual information for bivariate Gamma distributions. Image registration and change detection techniques based on bivariate gamma distributions are finally investigated. Simulation results conducted on synthetic and real data are very encouraging. Bivariate gamma distributions are good candidates allowing us to develop new image registration algorithms and new change detectors.

Index Terms— Multivariate gamma distributions, correlation coefficient, mutual information, maximum likelihood, image registration, image change detection.

I. INTRODUCTION

The univariate gamma distribution is uniquely defined in many statistical textbooks. However, extensions defining multivariate gamma distributions (MGDs) are more controversial. For instance, a full chapter of [1] is devoted to this problem (see also references therein). Most journal authors assume that a vector $\mathbf{x} = (x_1, \ldots, x_d)^T$ is distributed according to an MGD if the marginal distributions of x_i are univariate gamma distributions. However, the family of distributions satisfying this condition is very large. In order to reduce the size of the family of MGDs, S. Bar Lev and P. Bernardoff recently defined MGDs by the form of their moment generating function or Laplace transforms [2] [3]. The main contribution of this paper is to evaluate these distributions as candidates for image registration and change detection.

Given two remote sensing images of the same scene I, the reference, and J, the secondary image, the registration problem can be defined as follows: determine a geometric transformation T which maximizes the correlation coefficient between image I and the result of the transformation $T \circ J$. A fine modeling of the geometric deformation is required for the estimation of the coordinates of every pixel of the reference image inside the secondary image. The geometric deformation

is modeled by local rigid displacements [4]. The key element of the image registration problem is the estimation of the correlation coefficient between the images. This is usually done with an estimation window in the neighborhood of each pixel. In order to estimate the local rigid displacements with a good geometric resolution, one needs the smallest estimation window. However, this leads to estimations which may not be robust enough. In order to perform high quality estimations with a small number of samples, we propose to introduce a priori knowledge about the image statistics. In the case of power radar images, it is well known that the marginal distributions of pixels are gamma distributions [5]. Therefore, MGDs seem good candidates for the robust estimation of the correlation coefficient between radar images.

The change detection problem can be defined as follows. Consider two co-registered synthetic aperture radar (SAR) intensity images I and J acquired at two different dates t_I and t_J. Our objective is to produce a map representing the changes occurred in the scene between time t_I and time t_J. The final goal of a change detection analysis is to produce a binary map corresponding to the two classes: *change* and *no change*. The problem can be decomposed into two steps: 1) generation of a change image and 2) thresholding of the change image in order to produce the binary change map. The overall detection performance will depend on both, the quality of the change image and the quality of the thresholding. In this work, we choose to concentrate on the first step of the procedure, that is, the generation of an indicator of change for each pixel in the image. The change indicator can be obtained by computing the local correlation between both images, for each pixel position. For interesting approaches in the field of unsupervised change image thresholding, the reader can refer to the works of Bruzzone and Fernández Prieto [6], [7], Bruzzone and Serpico [8] and Bazi et al. [9]. The change indicator can also be useful by itself. Indeed, the end user of a change map often wants, not only the binary information given after thresholding, but also an indicator of the change amplitude. In order to evaluate the quality of a change image independently of the choice of the thresholding algorithm, one can study the evolution of the probability of detection as a function of the probability of false alarm, when a sequence of constant thresholds is used for the whole image. As in the image registration problem, a small estimation window is required in order to obtain a high resolution detector, that is,

a detector being able to identify changes with a small spatial extent. Again, the introduction of a priori knowledge through MGDs may improve the estimation accuracy when a small number of samples is used.

This paper is organized as follows. Section II recalls some important results on MGDs. Section III studies estimators of the unknown parameters of a bivariate gamma distribution (BGD). These estimators are based on the classical maximum likelihood method and method of moments. Section IV studies interesting properties of the mutual information for BGDs. The application to image registration and change detection is discussed in section V. Conclusions are finally reported in section VI.

II. MULTIVARIATE GAMMA DISTRIBUTIONS

A. Definitions

A polynomial $P(\mathbf{z})$ with respect to $\mathbf{z} = (z_1, \dots, z_d)$ is *affine* if the one variable polynomial $z_j \mapsto P(\mathbf{z})$ can be written $Az_j + B$ (for any $j = 1, \dots, d$), where A and B are polynomials with respect to the z_i's with $i \neq j$. A random vector $\mathbf{x} = (x_1, \dots, x_d)^T$ is distributed according to an MGD on $\mathbb{R}_+^d$ with shape parameter q and scale parameter P (denoted as $\mathbf{x} \sim \mathcal{G}(q, P)$) if its moment generating function (also called Laplace transform) is defined as follows [3]:

$$\psi_{\mathcal{G}(q,P)}(\mathbf{z}) = \mathsf{E}\left(e^{-\sum_{i=1}^d x_i z_i}\right) = [P(\mathbf{z})]^{-q}, \quad (1)$$

where $q \geq 0$ and P is an affine polynomial. It is important to note the following points:

- the affine polynomial P has to satisfy appropriate conditions including $P(0) = 1$. In the general case, determining necessary and sufficient conditions on the pair (q, P) such that $\mathcal{G}(q, P)$ exist is a difficult problem. The reader is invited to look at [3] for more details,
- by setting $z_j = 0$ for $j \neq i$ in (1), we obtain the Laplace transform of x_i, which is clearly a gamma distribution with shape parameter q and scale parameter p_i, where p_i is the coefficient of z_i in P.

A BGD corresponds to the particular case $d = 2$ and is defined by its moment generating function

$$\psi(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^{-q}, \quad (2)$$

with the following conditions

$$p_1 > 0, \; p_2 > 0, \; p_{12} > 0, \; p_1 p_2 - p_{12} \geq 0. \quad (3)$$

In the bi-dimensional case, (3) are necessary and sufficient conditions for (2) to be the moment generating function of a probability distribution defined on $[0, \infty[^2$. Note again that (2) implies that the marginal distributions of x_1 and x_2 are "gamma distributions" (denoted as $x_1 \sim \mathcal{G}(q, p_1)$ and $x_2 \sim \mathcal{G}(q, p_2)$) with the following densities:

$$f_{1D}(x_i) = \frac{x_i^{q-1}}{\Gamma(q) p_i^q} \exp\left(-\frac{x_i}{p_i}\right) \mathbb{I}_{\mathbb{R}_+}(x_i)$$

where $\mathbb{I}_{\mathbb{R}_+}(x_i)$ is the indicator function defined on $[0, \infty[$ ($\mathbb{I}_{\mathbb{R}_+}(x_i) = 1$ if $x_i \geq 0$, $\mathbb{I}_{\mathbb{R}_+}(x_i) = 0$ else), for $i \in \{1, 2\}$. Here $\Gamma(\cdot)$ is the usual gamma function defined in [10, p. 255].

B. Bivariate Gamma pdf

Obtaining tractable expressions for the probability density function (pdf) of a MGD defined by (1) is a challenging problem. However, in the bivariate case, the problem is much simpler. Straightforward computations allow to obtain the following density (see [1, p. 436] for a similar result)

$$f_{2D}(\mathbf{x}) = \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1} x_2^{q-1}}{p_{12}^q \Gamma(q)} f_q(cx_1 x_2) \mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}),$$

where $c = \frac{p_1 p_2 - p_{12}}{p_{12}^2}$ and $f_q(z)$ is defined as follows

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{k! \Gamma(q+k)}. \quad (4)$$

Note that $f_q(z)$ is related to the confluent hypergeometric function (see [1, p. 462]).

C. BGD Moments

The Taylor series expansion of the Laplace transform ψ can be written:

$$\psi(z_1, z_2) = \sum_{k,l \geq 0} \frac{(-1)^{k+l}}{k! l!} \mathsf{E}\left[x_1^k x_2^l\right] z_1^k z_2^l. \quad (5)$$

The moments of a BGD can be obtained by differentiating (5). For instance, the mean and variance of x_i (denoted $E[x_i]$ and var(x_i) respectively) can be expressed as follows

$$\mathsf{E}[x_i] = q p_i, \; \text{var}(x_i) = q p_i^2, \quad (6)$$

for $i = 1, 2$. Similarly, the covariance cov(x_1, x_2) and correlation coefficient $r(x_1, x_2)$ of a BGD can be easily computed:

$$\text{cov}(x_1, x_2) = \mathsf{E}[x_1 x_2] - \mathsf{E}[x_1]\mathsf{E}[x_2] = q(p_1 p_2 - p_{12}), \quad (7)$$

$$r(x_1, x_2) = \frac{\text{cov}(x_1, x_2)}{\sqrt{\text{var}(x_1)}\sqrt{\text{var}(x_2)}} = \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \quad (8)$$

It is important to note that for a known value of q, a BGD is fully characterized by $\boldsymbol{\theta} = (\mathsf{E}[x_1], \mathsf{E}[x_2], r(x_1, x_2))$ which will be denoted $\boldsymbol{\theta} = (m_1, m_2, r)$ in the remaining of the paper. Indeed, $\boldsymbol{\theta}$ and (p_1, p_2, p_{12}) are obviously related by a one-to-one transformation. Note also that the conditions (3) ensure that the covariance and correlation coefficient of the couple (x_1, x_2) are both positive.

More computations allow to obtain a general formula for the moments $\mathsf{E}[x_1^m x_2^n]$, for $(m, n) \in \mathbb{N}^2$, of a BGD:

$$\mathsf{E}[x_1^m x_2^n] = m_1^m m_2^n \frac{(q)_m}{q^m} \frac{(q)_n}{q^n} \sum_{k=0}^{\min(m,n)} \frac{(-m)_k (-n)_k}{(q)_k} \frac{r^k}{k!}, \quad (9)$$

where $(a)_k$ is the Pochhammer symbol defined by $(a)_0 = 1$ and

$$(a)_{k+1} = (a+k)(a)_k = a(a+1)\dots(a+k),$$

for any integer k (see [10, p. 256]). The mutual information of a BGD is related to the moments of $\sqrt{x_1 x_2}$ and $\log(x_j)$ for $j = 1, 2$. Straightforward computations detailed in appendices I and II yield the following results:

$$\mathsf{E}[\log(x_j)] = \psi(q) + \log\left(\frac{m_j}{q}\right), \quad (10)$$

and

$$\mathsf{E}\left[\sqrt{x_1 x_2}\right] = \frac{\sqrt{m_1 m_2}}{q} \frac{\Gamma(q+\frac{1}{2})^2}{\Gamma(q)^2} {}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; r\right), \quad (11)$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function and ${}_2F_1$ is the Gauss's hypergeometric function (see [10, p. 555–566]).

III. PARAMETER ESTIMATION

This section addresses the problem of estimating the unknown parameter vector $\boldsymbol{\theta}$ from n independent vectors $\mathbf{x} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$, where $\mathbf{x}^i = (x_1^i, x_2^i)$ is distributed according to a BGD with parameter vector $\boldsymbol{\theta}$. Note that the parameter q is assumed to be known here, as in most practical applications. However, this assumption could be relaxed.

A. Maximum Likelihood Method

1) Principles: The maximum likelihood (ML) method can be applied in the bivariate case ($d = 2$) since a closed-form expression of the density is available[1]. In this particular case, after removing the terms which do not depend on $\boldsymbol{\theta}$, the log-likelihood function can be written as follows:

$$l(\mathbf{x}; \boldsymbol{\theta}) = -nq \log(m_1 m_2) - \sum_{j=1}^{2} \frac{nq\overline{x}_j}{m_j(1-r)} \\ - nq \log(1-r) + \sum_{i=1}^{n} \log f_q(c x_1^i x_2^i), \quad (12)$$

where $c = \frac{rq^2}{m_1 m_2 (1-r)^2}$, and $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_j^i$ is the sample mean of x_j for $j = 1, 2$. By differentiating the log-likelihood with respect to m_1, m_2 and r, and by noting that $f_q'(z) = f_{q+1}(z)$, the following set of equations is obtained

$$\frac{nq\overline{x}_i}{1-r} - nqm_i - \frac{r}{(1-r)^2}q^2 m_i \Delta = 0, \quad i \in \{1, 2\}, \quad (13)$$

$$\frac{nq\overline{x}_1}{(1-r)m_1} + \frac{nq\overline{x}_2}{(1-r)m_2} - nq - \frac{1+r}{(1-r)^2}q^2 \Delta = 0, \quad (14)$$

where

$$\Delta = \frac{1}{m_1 m_2}\left(\sum_{i=1}^{n} x_1^i x_2^i \frac{f_{q+1}(c x_1^i x_2^i)}{f_q(c x_1^i x_2^i)}\right). \quad (15)$$

The maximum likelihood estimators (MLEs) of m_1 and m_2 are then easily obtained from these equations:

$$\widehat{m}_{1\text{ML}} = \overline{x}_1, \quad \widehat{m}_{2\text{ML}} = \overline{x}_2. \quad (16)$$

After replacing m_1 and m_2 by their MLEs in (14), we can easily show that the MLE of r is obtained by computing the root $r \in [0, 1[$ of the following function

$$g(r) = r - 1 + \frac{q}{n\overline{x}_1\overline{x}_2}\left(\sum_{i=1}^{n} x_1^i x_2^i \frac{f_{q+1}(\widehat{c} x_1^i x_2^i)}{f_q(\widehat{c} x_1^i x_2^i)}\right) = 0, \quad (17)$$

[1]The problem is much more complicated in the general case where $d > 2$ since there is no tractable expression for the MGD density. In this case, the coefficients of P can be estimated by maximizing an appropriate composite likelihood criterion such as the *pairwise log-likelihood*. The reader is invited to consult [11] for more details.

where

$$\widehat{c} = \frac{r}{(1-r)^2}\frac{q^2}{\overline{x}_1\overline{x}_2}.$$

This is achieved by using a Newton-Raphson procedure initialized by the standard correlation coefficient estimator (defined in (25)). The convergence of the Newton-Raphson procedure is generally obtained after few iterations.

2) Performance: The asymptotic properties of the ML estimators $\widehat{m}_{1\text{ML}}$ and $\widehat{m}_{2\text{ML}}$ can be easily derived from the moments of the univariate gamma distributions $\mathcal{G}(q, p_1)$ and $\mathcal{G}(q, p_2)$. These estimators are obviously unbiased, convergent and efficient. However, the performance of $\widehat{r}_{\text{ML}}$ is more difficult to study. Of course, the MLE is known to be asymptotically unbiased and asymptotically efficient, under mild regularity conditions. Thus, the mean square error (MSE) of the estimates can be approximated for large data records by the Cramer-Rao lower bound (CRLB). For unbiased estimators, the CRLBs of the unknown parameters m_1, m_2 and r can be computed by inverting the Fisher information matrix, whose elements are defined by

$$[I(\boldsymbol{\theta})]_{ij} = -\mathsf{E}\left[\frac{\partial^2 \log f_{2D}(\mathbf{x})}{\partial\theta_i\partial\theta_j}\right], \quad (18)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T = (m_1, m_2, r)^T$. However this computation is difficult because of the term $\log f_q$ appearing in the log-likelihood. In such situation, it is very usual to approximate the expectations by using Monte Carlo methods. More specifically, this approach consists of approximating the elements of the Fisher information matrix $I(\boldsymbol{\theta})$ as follows

$$[I(\boldsymbol{\theta})]_{ij} \simeq -\frac{1}{N}\sum_{k=1}^{N} \frac{\partial^2 \log f_{2D}(\mathbf{x}_k)}{\partial\theta_i\partial\theta_j}, \quad (19)$$

where $\mathbf{x}_k$ is distributed according to the BGD of density f_{2D} and N is the number of Monte Carlo runs.

B. Method of Moments

1) Principles: This section briefly recalls the principle of the method of moments. Consider a function $\mathbf{h}(\cdot) : \mathbb{R}^M \to \mathbb{R}^L$ and the statistic s_n of size L defined as:

$$\mathbf{s}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{h}(\mathbf{x}^i), \quad (20)$$

where $\mathbf{h}(\cdot)$ is usually chosen such that $\mathbf{s}_n$ is composed of empirical moments. Denote as:

$$\mathbf{f}(\boldsymbol{\theta}) = \mathsf{E}[\mathbf{s}_n] = \mathsf{E}[\mathbf{h}(\mathbf{x}^1)]. \quad (21)$$

The moment estimator of $\boldsymbol{\theta}$ is constructed as follows:

$$\widehat{\boldsymbol{\theta}}_{\text{Mo}} = \mathbf{g}(\mathbf{s}_n), \quad (22)$$

where $\mathbf{g}(\mathbf{f}(\boldsymbol{\theta})) = \boldsymbol{\theta}$. By considering the function

$$\mathbf{h}(\mathbf{x}) = (x_1, x_2, x_1^2, x_2^2, x_1 x_2),$$

the following result is obtained

$$\mathbf{f}(\boldsymbol{\theta}) = [m_1, m_2, m_1^2(1+q^{-1}), m_2^2(1+q^{-1}), m_1 m_2(1+rq^{-1})]. \quad (23)$$

The unknown parameters (m_1, m_2, r) can then be expressed as functions of $\mathbf{f}(\boldsymbol{\theta}) = (f_1, f_2, f_3, f_4, f_5)$. For instance, the following relations are obtained

$$m_1 = f_1, \quad m_2 = f_2, \quad r = \frac{f_5 - f_1 f_2}{\sqrt{(f_3 - f_1^2)(f_4 - f_2^2)}}. \quad (24)$$

yielding the standard estimators:

$$\widehat{m_1}_{\text{Mo}} = \overline{x}_1, \qquad \widehat{m_2}_{\text{Mo}} = \overline{x}_2,$$
$$\widehat{r}_{\text{Mo}} = \frac{\sum_{i=1}^{n}(x_1^i - \overline{x}_1)(x_2^i - \overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1^i - \overline{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2^i - \overline{x}_2)^2}}. \quad (25)$$

2) Performance: The asymptotic performance of the estimator $\hat{\boldsymbol{\theta}}_{\text{Mo}}$ can be derived by imitating the results of [12] derived in the context of time series analysis. A key point of these proofs is the assumption $\mathbf{s}_n \overset{a.s.}{\to} \mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ which is verified herein by applying the strong law of large numbers to (20). As a result, the asymptotic MSE of $\hat{\boldsymbol{\theta}}_{\text{Mo}}$ can be derived:

$$\lim_{n\to\infty} n\mathsf{E}[(\hat{\boldsymbol{\theta}}_{\text{Mo}} - \boldsymbol{\theta})^2] = \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta})^t, \quad (26)$$

where $\mathbf{G}(\boldsymbol{\theta})$ is the Jacobian matrix of the vector $\mathbf{g}(\cdot)$ at point $\mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \lim_{n\to\infty} n\mathsf{E}\left[(\mathbf{s}_n - \mathbf{s})(\mathbf{s}_n - \mathbf{s})^T\right]. \quad (27)$$

In the previous example, according to (24), $g : \mathbb{R}^5 \to \mathbb{R}^3$ is defined as follows

$$\mathbf{g}(\mathbf{x}) = \left(x_1, x_2, \frac{x_5 - x_1 x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}}\right). \quad (28)$$

The partial derivatives of g_1 and g_2 with respect to $x_i, i = 1, \ldots, 5$ are trivial. By denoting $\gamma = \sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}$, those of g_3 can be expressed as

$$\frac{\partial g_3}{\partial x_1} = -\frac{x_2}{\gamma} + \frac{x_1(x_4 - x_2^2)(x_5 - x_1 x_2)}{\gamma^3},$$
$$\frac{\partial g_3}{\partial x_2} = -\frac{x_1}{\gamma} + \frac{x_2(x_3 - x_1^2)(x_5 - x_1 x_2)}{\gamma^3},$$
$$\frac{\partial g_3}{\partial x_3} = \frac{(x_1 x_2 - x_5)(x_4 - x_2^2)}{2\gamma^3}, \quad (29)$$
$$\frac{\partial g_3}{\partial x_4} = \frac{(x_1 x_2 - x_5)(x_3 - x_1^2)}{2\gamma^3},$$
$$\frac{\partial g_3}{\partial x_5} = \frac{1}{\gamma}.$$

The elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be computed from the moments of $\mathbf{h}(\mathbf{x})$ which are obtained by differentiating the Laplace transform (2). The asymptotic MSEs (26) are then computed by using (9).

IV. MUTUAL INFORMATION FOR BGDs

Some limitations of the standard estimated correlation coefficient can be alleviated by using other similarity measures [4]. These similarity measures include the well-known mutual information. The mutual information of a BGD of shape parameter q and scale parameter $\mathbf{P} = (p_1, p_2, p_{12})$ can be defined as follows:

$$M_q(p_1, p_2, p_{12}) = \int_{\mathbb{R}^2} f_{2D}(\mathbf{x}) \log\left[\frac{f_{2D}(\mathbf{x})}{f(x_1, .)f(., x_2)}\right] d\mathbf{x}, \quad (30)$$

where $f(x_1, .)$ and $f(., x_2)$ are the marginal densities of the vector $\mathbf{x} = (x_1, x_2)$ and $f_{2D}(\mathbf{x})$ is its joint pdf. This section shows that the mutual information of BGDs is related to the correlation coefficient r by a one-to-one transformation. Interesting approximations of this mutual information for $r \to 0$ and $r \to 1$ are also derived.

A. Numerical Evaluation of the mutual information

By replacing the densities $f(x_1, .)$, $f(x_2, .)$ and $f_{2D}(\mathbf{x})$ by their analytical expressions, the following results can be obtained:

$$M_q(p_1, p_2, p_{12})$$
$$= q \log\left(\frac{p_1 p_2}{p_{12}}\right) - c\left(\mathsf{E}[x_1]\frac{p_{12}}{p_1} + \mathsf{E}[x_2]\frac{p_{12}}{p_2}\right) \quad (31)$$
$$+ \mathsf{E}\{\log[\Gamma(q)f_q(cx_1 x_2)]\}.$$

The first terms of $M_q(p_1, p_2, p_{12})$ can be easily expressed as a function of $\boldsymbol{\theta}$ by using the mean of a univariate gamma distribution given in (6). The mutual information $M_q(p_1, p_2, p_{12})$ can then be expressed as follows:

$$M_q(p_1, p_2, p_{12}) = q \log(1 - r) - \frac{2qr}{1 - r} \quad (32)$$
$$+ \mathsf{E}\{\log[\Gamma(q)f_q(cx_1 x_2)]\}.$$

However, a simple closed form expression for $A = \mathsf{E}\{\log[\Gamma(q)f_q(cx_1 x_2)]\}$ cannot be obtained, requiring to use a numerical procedure for its computation.

The numerical evaluation of A can be significantly simplified by noting that (x_1, x_2) and $(\alpha x_1, \beta x_2)$ have the same mutual information for any $(\alpha, \beta) \in \mathbb{R}^2$. Indeed, this property implies the following result:

$$M_q(p_1, p_2, p_{12}) = M_q\left(1, 1, \frac{p_{12}}{p_1 p_2}\right) = M_q(1, 1, 1 - r), \quad (33)$$

where $1 - r = \frac{p_{12}}{p_1 p_2} \in [0, 1]$. As a consequence, $A = \mathsf{E}\{\log[\Gamma(q)f_q(cx_1 x_2)]\}$ can be computed by replacing (p_1, p_2, p_{12}) by $(1, 1, 1 - r)$, where $1 - r \in [0, 1]$. This expectation can be pre-computed for all possible values of q and for $1 - r \in]0, 1]$, simplifying the numerical evaluation of M_q.

Moreover, it is interesting to note that (33) shows that the mutual information M_q and the correlation r are related by a one-to-one transformation. Consequently, M_q and r should provide similar performance for image registration and change detection. The advantage of using the mutual information will be discussed later.

B. Approximations of the mutual information

The numerical evaluation of A can be avoided for values of r closed to 0 and 1 by using approximations. Indeed, the following results can be obtained:

1) $r \to 0$: the second-order Taylor expansion of $f_q(z)$ around $z = 0$ can be written

$$f_q(z) = 1 + \frac{z}{q\Gamma(q)} + \frac{z^2}{2q(q+1)\Gamma(q)} + o(z^2), \quad (34)$$

where $o(z^2)/z^2$ tends to 0, as $z \to 0$. As a consequence, A can be approximated as follows:

$$A \approx \frac{c\mathsf{E}[x_1 x_2]}{q} - \frac{c^2 \mathsf{E}[(x_1 x_2)^2]}{2q} \left(\frac{1}{q(q+1)} \right). \quad (35)$$

By using (9), the mutual information M_q can be finally approximated as follows

$$M_q \approx \frac{r^2}{2}. \quad (36)$$

2) $r \to 1$: The Taylor expansion of $f_q(z)$ around ∞ can be written

$$f_q(z) = \frac{\exp(2\sqrt{z})}{\sqrt{4\pi z^{q-1/2}}} (1 + o(1)), \quad (37)$$

where $o(1)$ tends to 0, as $z \to \infty$. As a consequence, A can be approximated as follows:

$$\begin{aligned} A &\approx \mathsf{E} \left\{ \log \left[\frac{\exp(2\sqrt{cx_1 x_2})}{\sqrt{4\pi(cx_1 x_2)^{q-1/2}}} \right] \right\}, \\ &\approx 2\sqrt{c}\, \mathsf{E}[\sqrt{x_1 x_2}] - \left(\frac{q}{2} - 1 \right) \Big(\log c \\ &\quad + \mathsf{E}[\log x_1 + \log x_2] \Big) + \log \left[\frac{\Gamma(q)}{2\sqrt{\pi}} \right]. \end{aligned} \quad (38)$$

After replacing the means of $\sqrt{x_1 x_2}$, $\log(x_1)$ and $\log(x_2)$ derived in appendices I and II, the following result can be obtained

$$\begin{aligned} M_q &\approx -\frac{1}{2} \log(1 - r) + \left(q - \frac{1}{2} \right) + \log \left[\frac{\Gamma(p)}{2\sqrt{\pi}} \right] \\ &\quad - \left(q - \frac{1}{2} \right) \psi(q). \end{aligned} \quad (39)$$

Figure 1 shows that the mutual information M_q can be accurately approximated by (36) and (39) for $r < 0.5$ and $r > 0.9$. This figure has been obtained with the parameters $p_1 = 1$ and $p_2 = 1$ without loss of generality (see discussion at the beginning of this section).
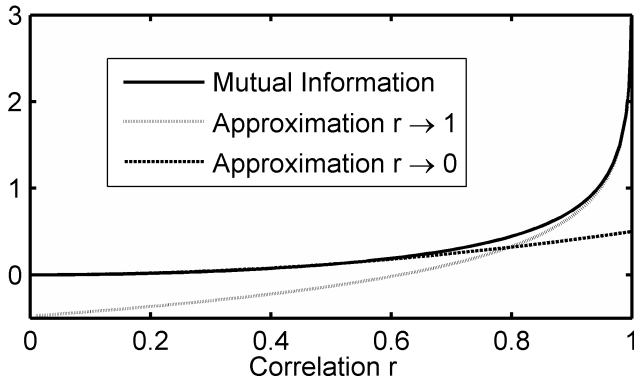

Fig. 1. Mutual information and its approximations for $r \to 0$ and $r \to 1$.

V. APPLICATION TO IMAGE REGISTRATION AND CHANGE DETECTION

This section explains carefully how BGDs can be used for image registration and change detection. Theoretical results are illustrated by many simulations conducted with synthetic and real data.

A. Synthetic Data

1) Generation: the generation of a vector $\mathbf{x} = (x_1, x_2)^T$ distributed according to a BGD has been performed as follows:
- simulate $2q$ independent multivariate Gaussian vectors of $\mathbb{R}^2$ denoted as $z^1, \ldots, z^{2q}$ with means $(0, 0)$ and the following 2×2 covariance matrix:

$$C = (c_{i,j})_{1 \le i,j \le 2} = \left(r^{\frac{|i-j|}{2}} \right)_{1 \le i,j \le 2},$$

- compute the kth component of $\mathbf{x} = (x_1, x_2)^T$ as $x_k = \frac{m_k}{2q} \sum_{1 \le i \le 2q} (z_k^i)^2$, where z_k^i is the kth component of z^i.

By computing the Laplace transform of $\mathbf{x}$, it can be shown that the two previous steps allow to generate random vectors $\mathbf{x} = (x_1, x_2)^T$ distributed according to a BGD whose marginal distributions are univariate gamma distributions $\mathcal{G}(q, m_1/q)$ and $\mathcal{G}(q, m_2/q)$. Moreover, the correlation coefficient of $\mathbf{x} = (x_1, x_2)^T$ is equal to r (the reader is invited to consult Appendix III for more details).

2) Estimation Performance: the first simulations compare the performance of the method of moments with the ML method as a function of n. Note that the possible values of n are $n = (2p+1) \times (2p+1)$, where $p \in \mathbb{N}$ (more precisely $3 \times 3 = 9$, $5 \times 5 = 25$, ..., $25 \times 25 = 625$). These values are appropriate for the image registration and change detection problems, as explained in the next sections. The number of Monte Carlo runs is 1000 for all figures presented in this section. The other parameters for this first example are $m_1 = 400$, $m_2 = 800$ and $q = 1$ (1-Look images). Figures 2 and 3 show the MSEs of the estimated correlation coefficient for two different correlation structures ($r = 0.2$ and $r = 0.8$). The circle curves correspond to the estimator of moments whereas the triangle curves correspond to the MLE. These figures show the interest of the ML method, which is much more efficient for this problem than the method of moments. The figures also show that the difference between the two methods is more significant for large values of the correlation coefficient r. Note that the theoretical asymptotic MSEs of both estimators determined in (18) and (26) are also displayed on Figs. 2 and 3 (continuous lines). The theoretical MSEs are clearly in good agreement with the estimated MSEs, even for small values of n. This is particularly true for large values of r.

3) Detection Performance: we consider synthetic vectors $\mathbf{x} = (x_1, x_2)^T$ (coming from 128×128 synthetic images) distributed according to BGDs with $r = 0.3$ and $r = 0.65$ modeling the presence and absence of changes, respectively. The correlation coefficient r of each bivariate vector $\mathbf{x}^{(i,j)} = (x_1^{(i,j)}, x_2^{(i,j)})^T$ (for $i, j = 1 \ldots 128$) is estimated from vectors belonging to windows of size $n = (2p+1) \times (2p+1)$ centered around the pixel of coordinates (i, j) in the two
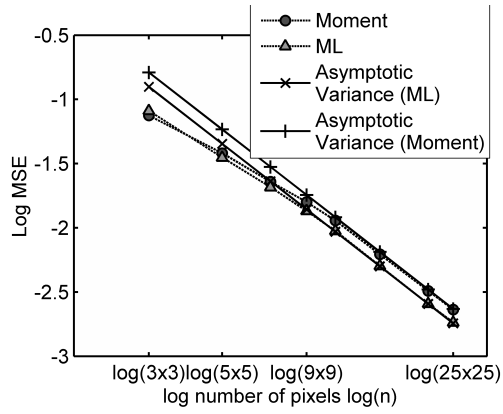
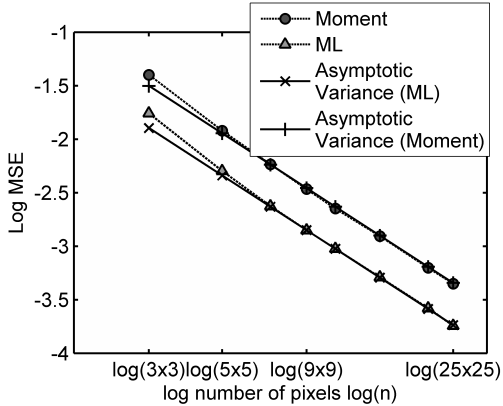Fig. 2. log MSEs versus $\log(n)$ for parameter r ($r = 0.2$).

Fig. 3. log MSEs versus $\log(n)$ for parameter r ($r = 0.8$).

analyzed images. The following binary hypothesis test is then considered:

$$
\begin{aligned}
H_0 \quad &\text{(absence of change)}: \quad \widehat{r} > \lambda, \\
H_1 \quad &\text{(presence of change)}: \quad \widehat{r} < \lambda,
\end{aligned}
\tag{40}
$$

where λ is a threshold depending on the probability of false alarm and $\widehat{r}$ is an estimator of the correlation coefficient (obtained from the method of moments or the maximum likelihood principle). The performance of the change detection strategy (40) can be defined by the two following probabilities [13, p. 34]

$$
\begin{aligned}
P_{\mathrm{D}} &= P\left[\text{accepting } \mathcal{H}_1 \,|\, \mathcal{H}_1 \text{ is true}\right] \\
&= P\left[\widehat{r} < \lambda \,|\, \mathcal{H}_1 \text{ is true}\right] = \int_{-\infty}^{\lambda} p_1(u)\,du,
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
P_{\mathrm{FA}} &= P\left[\text{accepting } \mathcal{H}_1 \,|\, \mathcal{H}_0 \text{ is true}\right] \\
&= P\left[\widehat{r} > \lambda \,|\, \mathcal{H}_1 \text{ is true}\right] = \int_{\lambda}^{\infty} p_0(u)\,du,
\end{aligned}
\tag{42}
$$

where $p_0(u)$ and $p_1(u)$ are the pdfs of $\widehat{r}$ under hypotheses H_0 and H_1, respectively. Thus, for each value of λ, there exists a pair $(P_{\mathrm{FA}}, P_{\mathrm{D}})$. The curves of P_D as a function of P_{FA} are called receiver operating characteristics (ROCs) [13, p. 38].
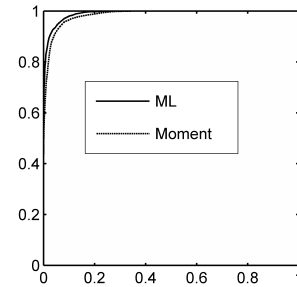
The ROCs for the change detection problem (40) are depicted on figures 4(a), 4(b) and 4(c) for three different window sizes corresponding to $n = 2p + 1 \in \{9, 15, 21\}$. The ML estimator clearly outperforms the moment estimator

for these examples. However, it is interesting to note that the two estimators have similar performances for large window sizes.
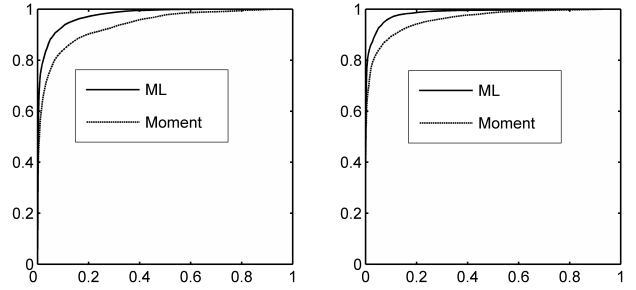
(a) $n = 9 \times 9$

(b) $n = 15 \times 15$

(c) $n = 21 \times 21$

Fig. 4. ROCs for synthetic data for different window sizes.

B. Application to Image Registration

This section studies an image registration technique based on BGDs. More precisely, consider two images whose pixels are denoted $\{x_1^1, \ldots, x_1^n\}$ and $\{x_2^1, \ldots, x_2^n\}$. Given the left image x_1, we propose the following basic 3-step image registration algorithm:

- Step 1: determine the search area in the right image x_2. Here, we use images that have been previously registered by a human operator using appropriate interactive software, a digital elevation model and geometrical sensor models. The use of registered images allows us to validate the results, since the expected shift between the images is equal to 0. For this experiment and without loss of generality, the search area is reduced to a line (composed of 10 pixels before and 10 pixels after the pixel of interest),
- Step 2: for each pixel x_2^j in the search area, estimate a similarity measure (correlation coefficient or mutual information) between x_1^i and x_2^j,
- Step 3: select the pixel providing the largest similarity.

This 3-step procedure has been applied to a couple of Radarsat 1-Look images acquired before and after the eruption of the

Nyiragongo volcano which occurred in January 2002. The Radarsat images are depicted on figures 5(a) (before eruption) and 5(b) (after eruption). Note that some changes due to the eruption can be clearly seen on the landing track for example. Figure 5(c) indicates the pixels of the image which have been affected by the eruption (white pixels). This reference map was obtained by photo-interpreters – who used the same SAR images we are using – and ground truth elaborated by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Humanitarian Information Center (HIC) on 27 January 2002, that is a few days after the eruption. This reference map was afterwards validated by a terrain mission. The types of change covered are: presence of a lava flow over old existing lava flows, damaged buildings (areas with different types of habitat). The area of study does not include forest or areas of dense vegetation (see `http://users.skynet.be/technaphot/webgomma/index.htm` for some examples of damages).
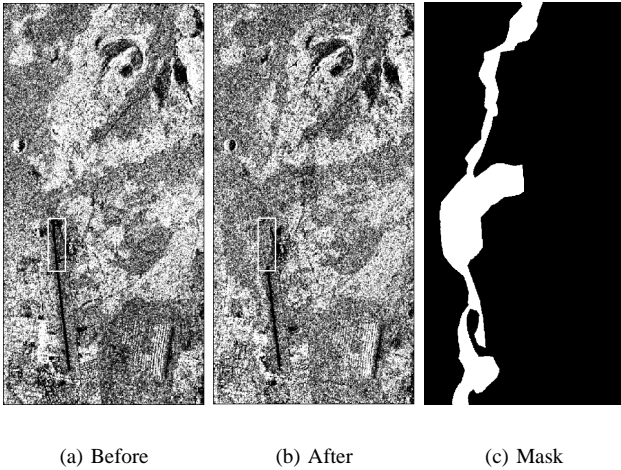

(a) Before (b) After (c) Mask

Fig. 5. Radarsat images of the Nyiragongo volcano.

Figures 6(a), 6(b) and 6(c) show an average of the estimated correlation coefficients with errorbars corresponding to mean $\pm$ standard deviation/10. These estimates have been computed for all black pixels which have not been affected by the eruption for different window sizes. More precisely, for every black pixel x_1^i of the left figure, we consider a window of size $n = (2p+1) \times (2p+1)$ centered around x_1^i. The same window is also considered in the right picture around pixel x_2^j. The correlation coefficient between the two pixels x_1^i and x_2^j is estimated by using the $n = (2p+1) \times (2p+1)$ couples of pixels located in the left and right windows. This operation is repeated for different central pixels x_2^j belonging to the search area (i.e. the 21 pixels of $[x_2^{i+\tau}]_{-10 \leq \tau \leq 10} = [x_2^{i-10}, \ldots, x_2^{i-1}, x_2^i, x_2^{i+1}, \ldots, x_2^{i+10}]$), where τ is the shift between the right and left windows. The results are averaged over all black pixels displayed in the mask 5(c). The estimated correlation coefficient is maximum when $j = i$, or equivalently $\tau = 0$, i.e. when the left and right windows are centered at the same location. This result indicates that the correlation coefficient can be efficiently used for image registration. Moreover, it is interesting to study how the estimator selectivity

(which can be defined as the relative amplitude of the peak compared to that of the plateau) varies from one estimator to another and depends on the window size. In particular, the ML estimator provides a slightly better selectivity than the estimator of moments. Note that the errobars are very similar for the two estimators. Even if the different methods provide similar results for image registration, it is important to note that the proposed framework allows one to define an interesting joint distribution for the vector (x_1^i, x_2^j). This distribution might be used for other tasks as, for instance, joint image segmentation and classification of both data sets.
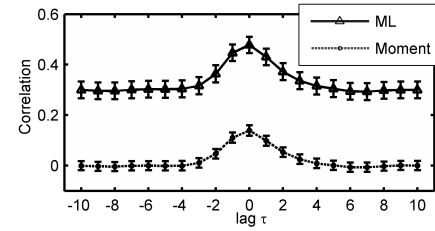
(a) Window Size $n = 7 \times 7$
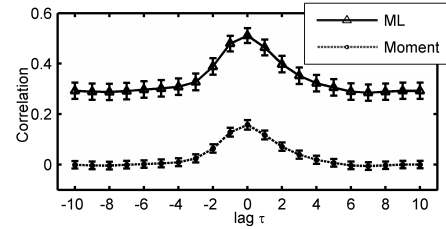
(b) Window Size $n = 9 \times 9$
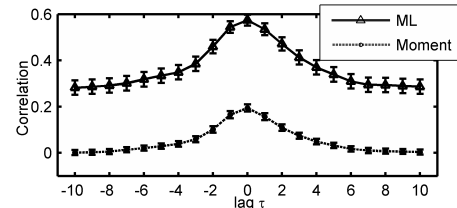
(c) Window Size $n = 15 \times 15$

Fig. 6. Averaged correlation coefficient estimates versus τ for black pixels with errorbars (ML: maximum likelihood estimator, Moment: moment estimator) for Nyiragongo images for several window sizes.

The same operation is conducted on a rectangular region composed of white pixels of the mask 5(c) (which have been affected by the eruption) depicted in white on figures 5(a) and 5(b). The results presented on Fig. 7 clearly show that the estimated correlation coefficient is much smaller when computed on a region affected by the eruption (and also that there is no peak which might be used for registration). This result is interesting and can be used for detecting changes between the two images, as illustrated in the next section.
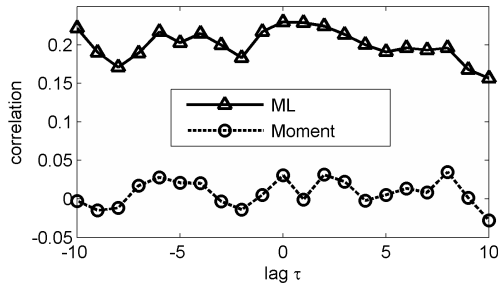
Fig. 7. Averaged correlation coefficient estimates versus τ for white pixels belonging to the Nyiragongo images square region (ML: maximum likelihood estimator, Moment: moment estimator).

C. Application to Change Detection

This section considers two 1-Look images acquired at different dates around Gloucester (England) before and during a flood (on Sept. 9, 2000 and Oct. 21, 2000 respectively). The images as well as a mask indicating the pixels affected by the flood are depicted on Figs. 8(a), 8(b) and 8(c). The reference map 8(c) was obtained by photo-interpreters – who used the same SAR images we are using – and a reference map built from Landsat and SPOT data acquired one day after the radar image.
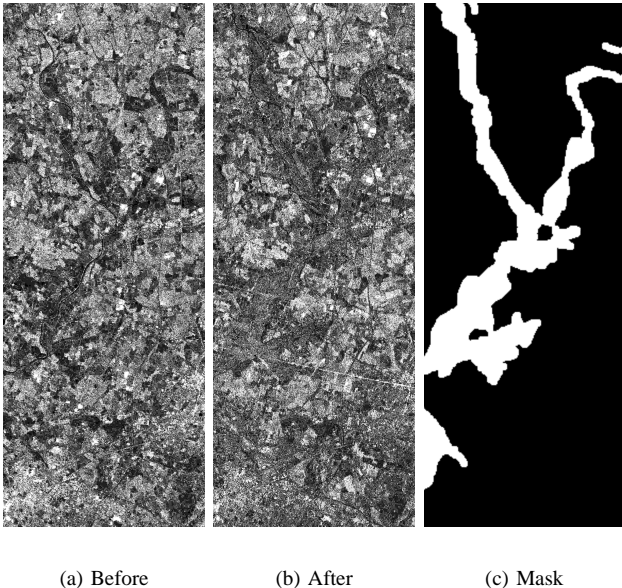

(a) Before   (b) After   (c) Mask

Fig. 8. Radarsat images of Gloucester before and after Flood.

This section compares the performance of the following change detectors

- the ratio edge detector which has been intensively used for SAR images [14], [15],
- the correlation change detector, where $\widehat{r}$ in (40) has been estimated with the moment estimator (referred to as "*Correlation Moment*"),
- the correlation change detector, where $\widehat{r}$ in (40) has been estimated with the ML method for BGDs (referred to as "*Correlation ML*").

The ROCs for this change detection problem are shown on figures 9(a), 9(b) and 9(c) for different window sizes n. The correlation ML detector clearly provides the best results.

(a) $n = 9 \times 9$ (b) $n = 15 \times 15$
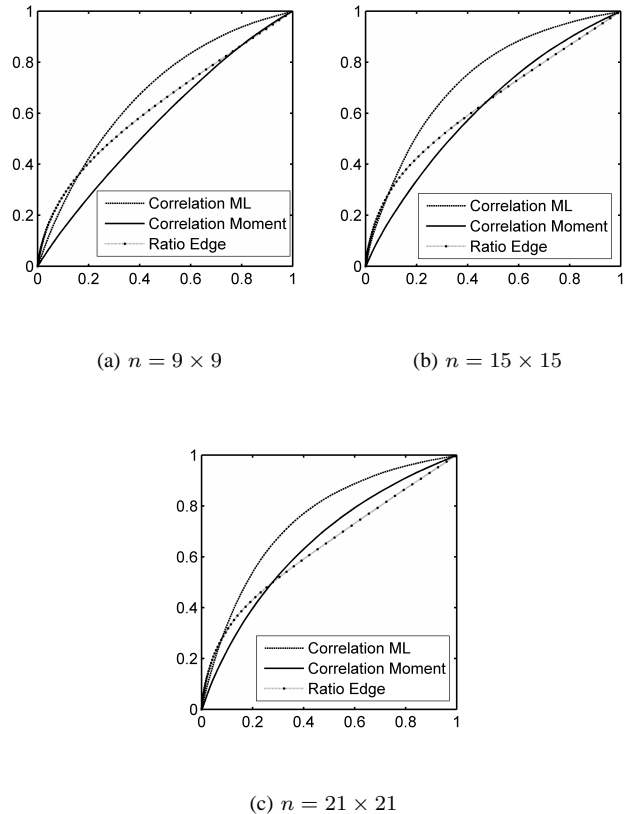

(c) $n = 21 \times 21$

Fig. 9. ROCs for Gloucester images for different window sizes.

The last experiments illustrate the advantage of using the mutual information for change detection. Consider the following change detector based on the mutual information:

$$\begin{aligned} H_0 \quad &(\text{absence of change}): \quad M_q(1,1,\widehat{r}) > \lambda_{PFA}, \\ H_1 \quad &(\text{presence of change}): \quad M_q(1,1,\widehat{r}) < \lambda_{PFA}, \end{aligned} \quad (43)$$

where $M_q(1,1,\widehat{r})$ is the estimated mutual information obtained by numerical integration of (32). The ROCs obtained with the detectors (40) and (43) are identical, reflecting the one-to-one transformation between the parameters $\widehat{r}$ and $M_q(1,1,\widehat{r})$. However, the advantage of using the mutual information for change detection is highlighted on Fig. 10, which shows the average probability of error $P_e = \frac{1}{2}(P_{ND} + P_{FA})$ (where $P_{ND} = 1 - P_D$ is the probability of non detection) as a function of the threshold λ for the change detectors (40) and (43). For a practical application, it is important to choose a threshold λ_{PFA} for these change detection problems. This choice can be governed by the value of the probability of error P_e. Assume that we are interested in having a probability of error satisfying $P_e \leq 0.39$. Figure 10 indicates that there are clearly more values of the threshold λ_{PFA} satisfying this condition for the curve "Mutual information" than for the curve "Correlation ML". This remains true whatever the value of the maximum probability of error P_e. Consequently, the

threshold is easier to be adjusted with the detector based on the mutual information (43) than the detector based on the correlation coefficient (40).
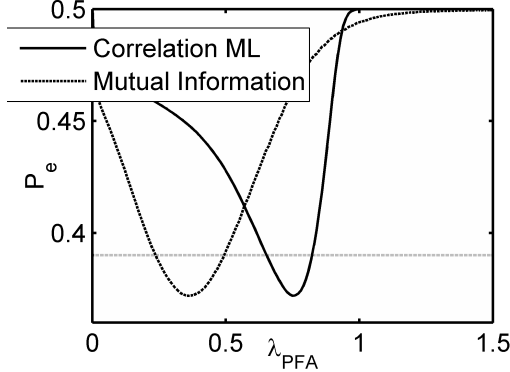

Fig. 10. Average probability of error $P_e = \frac{1}{2}(P_{FA} + P_{ND})$ versus threshold λ for Gloucester images for an estimation window of size $n = 9 \times 9$

VI. CONCLUSIONS

This paper studied the performance of image registration and change detection techniques based on bivariate gamma distributions. Both methods required to estimate the correlation coefficient between two images. Estimators based on the method of moments and on the maximum likelihood principle were studied. The asymptotic performance of both estimators was derived. The application to image registration and change detection was finally investigated.

The results showed the interest of using prior information about the data. On the other hand, the method presented here should not be used for more general cases where the BGD model does not hold. For these cases, the use of more general models as, for instance, copulas [16] or bivariate versions of the Pearson system [1, p. 6–9], should be studied.

VII. ACKNOWLEDGEMENTS

APPENDIX I
$E[\log U]$ WHERE $U \sim \Gamma(a, b)$

The moment of $\log U$ can be determined by the simple change of variable $V = bU$:

$$
\begin{aligned}
E[\log U] &= \int_0^\infty \log(u) \frac{b^a}{\Gamma(a)} e^{-bu} u^{a-1} du, \\
&= \frac{1}{\Gamma(a)} \int_0^\infty \log\left(\frac{v}{b}\right) e^{-v} v^{a-1} dv, \\
&= \frac{1}{\Gamma(a)} [\Gamma'(a) - \log(b)\Gamma(a)], \\
&= \psi(a) - \log(b). \quad (44)
\end{aligned}
$$

APPENDIX II
$E\left[\sqrt{x_1 x_2}\right]$ AND ITS APPROXIMATION FOR $r \to 1$ WHERE $(x_1, x_2) \sim \Gamma(q, \mathbf{P})$

A. $E\left[\sqrt{x_1 x_2}\right]$ computation

The moment of the random variable $\sqrt{x_1 x_2}$ is derived from the probability density function f_{2D} of the bivariate vector $\mathbf{x} = (x_1, x_2)^T$:

$$
\begin{aligned}
E\left[\sqrt{x_1 x_2}\right] &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \sqrt{x_1 x_2} f_{2D}(x_1, x_2) dx_1 dx_2, \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1/2} x_2^{q-1/2}}{p_{12}^q \Gamma(q)} \\
&\quad \times f_q(c x_1 x_2) \mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}).
\end{aligned}
$$

The definition of f_q given in (4) yields:

$$
\begin{aligned}
E\left[\sqrt{x_1 x_2}\right] &= \frac{1}{p_{12}^q \Gamma(q)} \sum_{k \geq 0} \frac{c^k}{k! \Gamma(q+k)} \\
&\quad \times \int_{\mathbb{R}_+} \exp\left(-\frac{p_2}{p_{12}} x_1\right) x_1^{q+k-1/2} dx_1 \\
&\quad \times \int_{\mathbb{R}_+} \exp\left(-\frac{p_1}{p_{12}} x_2\right) x_2^{q+k-1/2} dx_2, \\
&= \frac{1}{p_{12}^q \Gamma(q)} \sum_{k \geq 0} \frac{c^k}{k! \Gamma(q+k)} \left(\frac{p_{12}}{p_2}\right)^{q+k+1/2} \\
&\quad \times \Gamma(q+k+1/2) \left(\frac{p_{12}}{p_1}\right)^{q+k+1/2} \Gamma(q+k+1/2), \\
&= \left(\frac{p_{12}}{p_1 p_2}\right)^{q+1} \frac{\sqrt{p_1 p_2}}{\Gamma(q)} \sum_{k \geq 0} \frac{\Gamma(q+k+1/2)^2}{k! \Gamma(q+k)} r^k, \\
&= (1-r)^{q+1} \frac{\sqrt{m_1 m_2}}{q} \left(\frac{\Gamma(q+1/2)}{\Gamma(q)}\right)^2 \\
&\quad \times {}_2F_1\left(q+\frac{1}{2}, q+\frac{1}{2}; q; r\right),
\end{aligned}
$$

since

$$
r = \frac{p_1 p_2 - p_{12}}{p_1 p_2} = \frac{p_{12}^2}{p_1 p_2} c.
$$

Here ${}_2F_1$ is the Gauss's hypergeometric function (see [10, p. 555–566]) defined as:

$$
{}_2F_1(a, b; c; z) = \sum_{k=0}^\infty \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!},
$$

and $(a)_k$ is the Pochlammer symbol presented in section II-C (note that $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$ for any integer k and any real $a > 0$). By using the following properties of Gauss's hypergeometric functions:

1) The hypergeometric series ${}_2F_1(a, b; c; z)$ converges if c is not a negative integer for complex numbers z such that $|z| < 1$ or $|z| = 1$ if $\Re[c-a-b] > 0$.
2) ${}_2F_1(a, b; c; z) = (1-z)^{c-a-b} {}_2F_1(c-a, c-b; c; z)$ for all of $|z| < 1$ (see [10, p. 559]),

the following results can be obtained:

$$
E\left[\sqrt{x_1 x_2}\right] = \frac{\sqrt{m_1 m_2}}{q} \frac{\Gamma(q+1/2)^2}{\Gamma(q)^2} \, {}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; r\right).
$$

B. $E[\sqrt{x_1 x_2}]$ approximation for $r \to 1$

The following identity (Gauss's hypergeometric theorem):

$$_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)},$$

leads to:

$$_2F_1\left(\frac{-1}{2},\frac{-1}{2};q;1\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2},$$

and to the following first order Taylor expansion around $z = 1$:

$$_2F_1\left(\frac{-1}{2},\frac{-1}{2};q;z\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} + (z-1)$$
$$\times\ _2F_1'\left(\frac{-1}{2},\frac{-1}{2};q;1\right) + o(1-z),$$

where $o(1-z)/(1-z)$ tends to 0, as $z \to 1$. Using

$$_2F_1'(a,b;c;z) = \frac{ab}{c}\ _2F_1(a+1,b+1;c+1;z)\ \text{ for } |z| < 1,$$

the previous Taylor expansion can be written:

$$_2F_1\left(\frac{-1}{2},\frac{-1}{2};q;z\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2}\left(1+\frac{z-1}{4q}\right)$$
$$+ o(1-z).$$

Finally,

$$E\left[\sqrt{x_1 x_2}\right] = \frac{\sqrt{m_1 m_2}}{q}\frac{\Gamma(q+1/2)^2}{\Gamma(q)^2}\frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2}\left(1-\frac{1-r}{4q}\right)$$
$$+ o(1-r),$$
$$= \sqrt{m_1 m_2}\left(1-\frac{1-r}{4q}\right) + o(1-r)\quad \text{for } r \to 1.$$

APPENDIX III
GENERATION OF SYNTHETIC DATA DISTRIBUTED ACCORDING TO BGDs

This appendix shows that the vector $\mathbf{x} = (x_1, x_2)^T$ where $x_k = \frac{m_k}{2q}\sum_{1\le i\le 2q}(z_k^i)^2$ (where z_k^i is the kth component of $\mathbf{z}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, with $\mathbf{C} = (c_{i,j})_{1\le i,j\le 2} = \left(r^{\frac{|i-j|}{2}}\right)_{1\le i,j\le 2}$) is distributed according to a BGD whose marginals are Gamma distributions $\Gamma(q, m_1/q)$ and $\Gamma(q, m_2/q)$ and whose correlation coefficient is r. By using the independence between vectors $z^1, \ldots, z^{2q}$, the Laplace transform of $\mathbf{x}$ evaluated at $\mathbf{t} = (t_1, t_2)^T$ can be written:

$$E\left[\exp\left(-\mathbf{t}^T\mathbf{x}\right)\right] = E\left[\prod_{k=1}^2 \exp\left(-t_k x_k\right)\right],$$
$$= \prod_{i=1}^{2q} E\left[\exp\left(-\frac{(\mathbf{z}^i)^T\mathbf{S}(\mathbf{z}^i)}{2}\right)\right],$$

where

$$\mathbf{S} = \begin{pmatrix} t_1\frac{m_1}{q} & 0 \\ 0 & t_2\frac{m_2}{q} \end{pmatrix}.$$

By using the probability density function of a bivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C})$, the Laplace transform can be finally expressed as:

$$E\left[\exp\left(-\mathbf{t}^T\mathbf{x}\right)\right] = \prod_{i=1}^{2q}\int_{\mathbb{R}^2}\frac{\exp\left(-\mathbf{z}^T\left(\mathbf{C}^{-1}+\mathbf{S}\right)\mathbf{z}\right)}{(2\pi)\sqrt{\det \mathbf{C}}}d\mathbf{z},$$
$$= \left[\frac{1}{\det\left(\mathbf{I}_2+\mathbf{CS}\right)}\right]^q,$$
$$= \left[1 + \sum_{i=1}^2 \frac{m_i}{q}t_i + (1-r)\frac{m_1 m_2}{q^2}t_1 t_2\right]^{-q},$$

where $\mathbf{I}_2$ is the identity matrix in dimension 2. According to the definition (2), the vector $\mathbf{x}$ is distributed according to a BGD with shape parameter q and scale parameters $p_1 = \frac{m_1}{q}$, $p_2 = \frac{m_2}{q}$, $p_{12} = (1-r)\frac{m_1 m_2}{q^2}$. Property (8) ensures that the correlation coefficient of (x_1, x_2) is r.

REFERENCES

[1] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, vol. 1. New York: Wiley, 2nd ed., 2000.

[2] S. B. Lev, D. Bshouty, P. Enis, G. Letac, I. L. Lu, and D. Richards, "The diagonal natural exponential families and their classification," *J. of Theoret. Prob.*, vol. 7, pp. 883–928, 1994.

[3] P. Bernardoff, "Which multivariate Gamma distributions are infinitely divisible?," *Bernoulli*, 2006.

[4] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, pp. 2104–2120, Oct. 2004.

[5] T. F. Bush and F. T. Ulaby, "Fading characteristics of panchromatic radar backscatter from selected agricultural targets," *NASA STI/Recon Technical Report N*, vol. 75, pp. 18460–+, Dec. 1973.

[6] L. Bruzzone and D. F. Prieto, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 1171–1182, May 2000.

[7] L. Bruzzone and D. F. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *International Journal of Remote Sensing*, vol. 21, no. 4, pp. 817–822, 2000.

[8] L. Bruzzone and S. Serpico, "An Iterative Technique for the Detection of Land-Cover Transitions in Multitemporal Remote-Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 858–867, July 1997.

[9] Y. Bazi, L. Bruzzone, and F. Melgani, "An Unsupervised Approach Based on the Generalized Gaussian Model to Automatic Change Detection in Multitemporal SAR Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, pp. 874–887, Apr. 2005.

[10] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1972.

[11] F. Chatelain and J.-Y. Tourneret, "Composite likelihood estimation for multivariate mixed Poisson distributions," in *Proc. IEEE-SP Workshop Stat. Signal Processing*, (Bordeaux, France), July 2005.

[12] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *International Journal of adaptive control and signal processing*, vol. 3, pp. 191–229, 1989.

[13] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: Wiley, 1968.

[14] R. Touzi, A. Lopés, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, pp. 764–773, Nov. 1988.

[15] E. J. M. Rignot and J. J. van Zyl, "Change Detection Techniques for ERS-1 SAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, pp. 896–906, July 1993.

[16] R. B. Nelsen, *An Introduction to Copulas*, vol. 139 of *Lecture Notes in Statistics*. New York: Springer-Verlag, 1999.