

Aurora Classification Algorithm Based on Adaptive Lifting Wavelet Transform and LBP

Weibo Xing, Xuan Wang

School of Physics and Information Technology, Shaanxi Normal University, Xi'an Shaanxi
Email: xingweibo_0625@163.com

Received: May 2nd, 2016; accepted: May 22nd, 2016; published: May 25th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper presents a new dual-scaling algorithm based on adaptive lifting wavelet transform and improved Local Binary Pattern and classification of a combination of fuzzy neighbor Aurora classification algorithm. Based on the aurora image preprocessing, the algorithm is first using adaptive lifting wavelet transform of the original image to divide into several sub-images of Aurora, and then for each sub-image variable scale Gaussian filter, and to conduct sub-picture with the local binary pattern feature extraction, and finally with fuzzy neighbor classification algorithm to classify. Simulation results show that, first, the algorithm classification efficiency is higher than other Aurora classification algorithm, followed by the algorithm for ordinary noise, such as Gaussian noise and salt and pepper noise having better robustness.

Keywords

Adaptive Lifting Wavelet Transform, Two-Scale Algorithm, Local Binary Patterns, Fuzzy Nearest Neighbor Classifier

基于自适应提升小波变换和LBP的 极光分类算法

邢伟博, 王 晔

陕西师范大学物理学与信息技术学院, 陕西 西安
Email: xingweibo_0625@163.com

收稿日期: 2016年5月2日; 录用日期: 2016年5月22日; 发布日期: 2016年5月25日

摘要

本文提出了一种新的基于自适应提升小波变换的双尺度算法、改进的局部二值模式和模糊近邻分类相结合的极光分类算法。该算法在极光图像预处理的基础之上, 先是利用自适应提升的小波变换将原始的极光图像分为几个子图像, 然后再对各个子图像进行变尺度的高斯滤波。用局部二值模式进行对子图像进行特征的提取, 最后用模糊的近邻分类算法对其进行分类。仿真实验证明, 首先本文算法的分类效率高与其他极光分类算法, 其次就是本文算法对普通的噪声, 例如高斯噪声和椒盐噪声, 都有较好的鲁棒性。

关键词

自适应提升小波变换, 双尺度算法, 局部二值模式, 模糊近邻分类

1. 引言

极光隐藏着巨大的科学研究价值。从空间物理学的角度, 极光现象反映了太阳与地球之间某种动力学的关系。从天气物理的角度来讲, 极光不仅出现在地球上, 据观测表明, 太阳系中的某些星球也出现过极光, 所以研究极光对我们探索太阳系中的奥秘奠定了基础。从能量能源的角度来讲, 每爆发一次极光, 所产生的能量基本上都大于地球一年发电量总和。研究极光的意义并不仅限于此。所以对极光图像进行分类研究, 能帮助研究者们更加清晰的理解极光现象对我们生活的影响。

极光的研究起源很早, 早在 1995 年, Carl Stormer [1]通过分析大量的极光数据, 首次对极光图像提出了三种分类, 分为无放射状的结构、有放射状的机构和火焰形状的三种类别, 但是这是基于人工标记的办法, 分类效率不高并且基本没有鲁棒性。2004 年 Syrjäsoo 和他的研究团队[2]首次将数字图像处理和机器视觉技术引入到了极光分类的研究领域, 并且提出了自动的极光分类方法。虽然在分类的效率方面有了提高, 但是在拍摄过程中因为干扰都会产生的一些纹理信息丢失。2007 年, Wang [3]通过使用主成分分析法, 提出了一种基于表象特征的极光图像分类方法。2008 年, Gao 等人[4]为了降低极光图像的特征冗余, 提取基于极光图像的局部 Gabor 滤波器特征, 并且与动态过程结合的办法进行极光图像的分类, 取得了较高的分类准确率。到 2009 年, 胡泽俊博士另辟蹊径, 引入形态学的概念, 将极光图像分为了冕状形态极光和弧光状的极光两大类, 然后对冕状的再进行细分, 即热点冕状极光、辐射冕状极光和帷幔型的冕状三种极光。2011 年, Fu 等人[5]也将形态学分量分析(MCA)的方法用于极光图像的分类, 完成了对弧状和冕状两类极光图像的分类。这种方法在对极光图像进行 MAC 分离后, 再对其现有的纹理提取特征信息, 然后再分类, 提高了弧状和冕状极光的分类效率。在此基础之上, 又出现了一些基于监督的极光分类方法, 例如 BP 神经网络方法和支持向量机(SVM)的方法。虽然这些极光分类方法的效率在不断的提高, 但是并不能取得令人满意的效果, 最主要的问题就是极光图像变化快, 数据量非常大, 所以现在的问题就是如何针对海量的极光图像数据并且能快速而精确的分类出各类极光图像。

本文针对海量的极光数据, 提出了双尺度算法和局部二值模式以及近邻分类相结合的极光分类算法, 实验证明, 不仅分类效果较高, 而且其鲁棒性也非常好。

2. 自适应提升的小波变换

假设有一种最原始的信号 $S: Z^d \rightarrow R$ 表示的是首先用一种特殊的小波变换或者多向分解技术将信号 S 分解成一个近似信号 x 和细节描述信号 y 。然后用映射 U 代替 y 去改写 x , 然后得到一个新的近似信号

x' :

$$x' = x + U(y) \quad (1)$$

随后, 预测映射 P 用 x' 去改写 y , 然后产生一个新的信号 y' :

$$y' = y - P(x') \quad (2)$$

在信号的合成阶段, 通过反转提升步骤来重构原始信号。有保证的重构并且在提升的步骤中不需要一些特定的条件在图像的重构中是非常重要的。它灵活地为开发各种应用提供了可能性。

在早期的研究中, 为了对特殊图像的特征纹理处理并且在扭曲的图像中提取图像的有用信息, 使得自适应小波变换的方法有了快速的发展。

我们的自适应小波变换方法如下[6]所示, 将 (M, N) 阶的原始图像 $f(i, j)$ 用简单的多相分解技术分解成 4 个 $(M/2, N/2)$ 的子图像:

$$\begin{aligned} x(n, m) &= f(2n, 2m), y_h(n, m) = f(2n+1, 2m), \\ y_v(n, m) &= f(2n, 2m+1), y_d(n, m) = f(2n+1, 2m+1) \end{aligned} \quad (3)$$

然后, 输出的近似信号 $x(n, m)$ 等于:

$$x(n, m) = x(n, m) \oplus_{p_{nm}} U_{p_{nm}}(y_h(n, m), y_v(n, m), y_d(n, m)) \quad (4)$$

其中 p_{nm} 是二值映象的决策因子 D 的判决结果, 即

$$p_{nm} = D(n, m) = \begin{cases} 0 & \partial(n, m) > T \\ 1 & \partial(n, m) \leq T \end{cases} \quad (5)$$

T 是给出的阈值, $\partial(n, m)$ 的定义如下:

$$\begin{aligned} \partial(n, m) &= |y_h(n-1, m) - \mu| + |y_h(n, m) - \mu| + |y_v(n, m-1) - \mu| + |y_v(n, m) - \mu| \\ &\quad + |y_d(n-1, m) - \mu| + |y_d(n, m) - \mu| + |y_h(n, m-1) - \mu| + |y_h(n-1, m-1) - \mu| \end{aligned} \quad (6)$$

μ 是 $x(n, m)$ 表示的是水平方向、垂直和对角线方向的八种像素的平均值, 公式如下

$$\begin{aligned} \mu &= \frac{1}{8} [y_h(n-1, m) + y_h(n, m) + y_v(n, m-1) + y_v(n, m) \\ &\quad + y_d(n-1, m) + y_d(n, m) + y_d(n, m-1) + y_d(n-1, m-1)] \end{aligned} \quad (7)$$

在新的自适应提升的方法中, 随着 p_{nm} 的不同, 更新的步骤也是不同, 对每一种不同的选择, 有不同的添加操作 $\oplus_{p_{nm}}$ 和不同的更新操作 $U_{p_{nm}}$, 也就是说, 它是依靠位置并且能自适应的。更新因子 $U_{p_{nm}}$ 和加操作 $\oplus_{p_{nm}}$ 由如下定义

$$U_{p_{nm}} = \begin{cases} 0 & P_{nm} = 0 \\ \frac{1}{2} [y_h(n-1, m) + y_h(n, m) + y_v(n, m-1) + y_v(n, m)] & P_{nm} = 1 \end{cases} \quad (8)$$

$$x \oplus_{p_{nm}} u = \begin{cases} x + u & p_{nm} = 0 \\ \frac{1}{2}(x + u) & p_{nm} = 1 \end{cases} \quad (9)$$

三个预测操作 $P_h(x'(n, m))$, $P_v(x'(n, m))$ 以及 $P_d(x'(n, m))$ 满足下面的三个式:

$$P_h(x'(n, m)) = \frac{1}{2} [x'(n-1, m) + x'(n, m)] \quad (10)$$

$$P_v(x'(n, m)) = \frac{1}{2} [x'(n, m-1) + x'(n, m)] \quad (11)$$

$$P_d(x'(n, m)) = \frac{1}{2} [x'(n-1, m-1) + x'(n, m)] \quad (12)$$

息 $y'_h(n, m)$, $y'_v(n, m)$ 和 $y'_d(n, m)$ 由下面的方程计算:

$$\begin{aligned} y'_h(n, m) &= y_h(n, m) - P_h(x'(n, m)) \\ y'_v(n, m) &= y_v(n, m) - P_v(x'(n, m)) \\ y'_d(n, m) &= y_d(n, m) - P_d(x'(n, m)) \end{aligned} \quad (13)$$

自适应方案的说明如下: 在平滑区域当 $p_{mn} = 1$ 时, 用 $x(n, m)$ 和它的水平, 垂直, 对角线的加权均值来计算 $x(n, m)$, 而对于不平滑区域, 当 $p_{mn} = 0$, 不执行任何的滤波, 例如 $x'(n, m) = x(n, m)$, 这样做的主要原因是极在极光识别中, 不连续的图像在图像识别的过程中有很重要的意义, 但是必须要滤除很多孤立而且不连续的点。通过改变阈值 T , 一是可以平衡和过滤有用的信息。另外我们的自适应方案只包括加法和减法操作在空间域, 并持有高平滑的性能, 就地操作, 所以它提供了非常低的计算成本和简单的硬件实现。自适应小波变换的框图如图 1。

3. 本文提出的方法

本文的算法是首先用自适应提升的小波变换, 使得极光的特征更加的明显, 用变尺度的高斯滤波对各个子图像进行滤波, 使得图像的干扰大大降低, 然后使用局部二值模式进行累积成特征直方图, 算法的最后用模糊近邻分类方法对极光图像进行分类。

3.1. 双尺度分解算法

所以本文所提出的双尺度的图像分解由两个阶段组成, 在第一阶段中, 考虑到图像的不连续性, 纹理和轮廓在图像的识别中起着非常重的作用, 因为自适应的小波变换对轮廓和纹理的提取效果明显, 所以先用自适应提升的小波变换把原图像 $f(x, y)$ 分解成几个子图像 $f_s(x, y)$, 然后通过对图像畸变点的滤除得到有用的特征信息。在第二阶段中, 主要是为了提高算法对噪音的容忍限度, 因此需要对各个子图像进行变尺度的高斯滤波:

$$L_{s, \sigma}(x, y) = f_s(x, y) * G(x, y, \sigma) \quad (14)$$

在上面的公式中*代表的是卷积运算, 其中

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right) \quad (15)$$

3.2. 图像的特征提取

因为像素之间的距离是有相关性的, 所以主要的纹理信息可以从周围的像素中提取。因此, 在 DS-LBP 中, 仅仅需要考虑中心和周围的八个像素, 所以用下面这个表达式进行特征的提取

$$DS-LBP_{s, \sigma}(x, y) = s(DS_c - \mu)28 + \min_k \sum_{p=k}^{(7+k) \bmod 8} s(DS_p - \mu)2^p, \quad k = 0, \dots, 7 \quad (16)$$

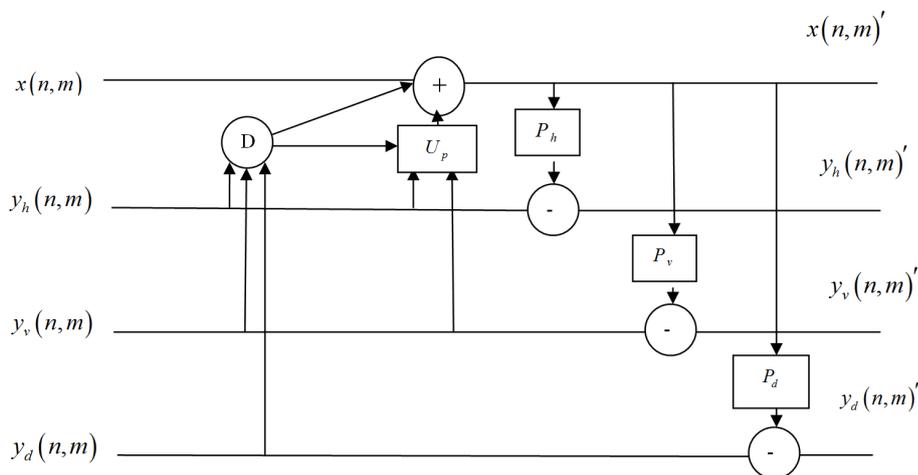


Figure 1. The block diagram of adaptive wavelet transform
图 1. 自适应小波变换框图

对于每一个图像 $L_{s,\sigma}(x,y)$ ，用(2-32)能计算出它的离散直方图， $0 \leq n < K$ ，and $K = 2^9$ ，将极光图像的特征用这种离散的归一化直方图的方式来表示。因为极光图像的位置与离散直方图不相关，所以提取的特征是不变的。

输入极光图像被转化成灰度级图像，接着根据点和中心最大的距离进行归一化，再根据双尺度图像分解将灰度级图像分解成一系列的子图像，在每个子图像中使用局部二值模式并且累积成离散的直方图 $DS-LBPHs,\sigma(n)$ ，表达式如下：

$$NDS-LBPHs,\sigma(n) = DS-LBPHs,\sigma(n) / N \times M \tag{17}$$

其中 $N \times M$ 是没个子图像的像素，这些归一化直方图被视为特性，再使用模糊的邻近分类进行分类。

3.3. 模糊的邻近分类器

在提出的模糊邻近分类器，两种极光的相似用以下的公式定义

$$J = \max_s \left(\frac{1}{N_\sigma} \sum_\sigma J_{s,\sigma} \right) \tag{18}$$

上式子中的 N_σ 表示的是高斯滤波的变尺度系数， $J_{s,\sigma}$ 表示的是在同一尺度和相同波段的模糊相似度，定义如下：

$$J_{s,\sigma} = \sum_{n=0}^{K-1} \sqrt{NDS-LBPH_{s,\sigma}(n) NDS-LBPH'_{s,\sigma}(n)} \tag{19}$$

这里 $K = 2^9$ ， $NDS-LBPH_{s,\sigma}(n)$ 和 $NDS-LBPH'_{s,\sigma}(n)$ 分别代表了具有相同的尺度和波段的两种极光的DS-LBP归一化直方图。我们很容易就发现 $J_{s,\sigma} \in [0,1]$ ，如果 $NDS-LBPH'_{s,\sigma}(n)$ 和 $NDS-LBPH_{s,\sigma}(n)$ 是非常的像，

即 $NDS-LBPH_{s,\sigma}(n) \approx NDS-LBPH'_{s,\sigma}(n)$ 则说明 $J_{s,\sigma} \approx 1$ ；同样的道理，假如 $J_{s,\sigma} \approx 0$ ，说明 $NDS-LBPH'_{s,\sigma}(n)$ 和 $NDS-LBPH_{s,\sigma}(n)$ 的重叠部分几乎没有，所以说这两种图像不是一类。

模糊邻近算法的模糊分类器主要包含以下三个步骤：

第一步：给出一张极光图像 x 和一个训练集 $S = \{y_i, 1 \leq i \leq L\}$ 我们用公式(16)计算 x 和 y 的模糊相似度，然后在测试集中选取和训练集最近邻的图像 $\{y_n, 1 \leq i \leq k\}$

第二步：相加之后，如果属于同一类 w 的 k 阶近似相似度，就用以下的公式。

$$S_{w_i} = \sum_{y_{n_j} \in w_i} J(x, y_{n_j})$$

第三步：计算最大的 S_{w_i} 并且对 x 进行归类。

4. 仿真实验

我们实验的数据是从极光的总数据库中随机的选取 3200 个样本进行测试, 其中 arc 的样本数量为 800 个, hospot 的极光样本数量为 800 个, 冕状的极光样本总数为 800 个, 小尺度状的极光样本数量为 800 个。首先在分类好的极光类中选取训练集, 在每一类中随机的选取 5 张极光图像作为训练样本, 然后在分类好的极光类中, 选取测试样本, 然后对选取的这些个测试集进行极光分类实验, 最后检验分类的正确性。

本章主要工作分为以下三个方面, 首先, 测试提出方法的分类的精确性, 并且和已经提出的方法的结果进行对比, 其次是验证其加噪声之后, 方法对噪声的抗噪程度, 即就是验证其鲁棒性。最后是测试算法的成本。

4.1. 极光分类算法效率对比

实验 1: 比较各类极光识别算法的分类效率(%)

表 1 即为比较各种分类算法而得到实验数据, 从表 1 可以看出, 本文提出的分类的算法, 在同样的实验条件之下, 得到的实验结果相对能比以上的各种都略好。

4.2. 算法鲁棒性

在这部分中, 为了验证所提出的方案的抗噪性能, 在极光图像中加入两种常见的噪声, 第一种就是常见的高斯噪声, 其均值为 0, 方差为 0.01。第二种就是采用密度为 0.05 的椒盐噪声。再用以上的方案进行实验, 然后计算本文的算法对噪声容量。

1) 在极光图像经过预处理之后, 在所有的测试样本中加入均值为 0. 方差为 0.01 的高斯噪声, 然后测试样本的分类效率(见表 2)。

2) 在原始的极光图像经过了预处理之后, 在测试样本中加入密度值是 0.05 的椒盐噪声, 然后测试其分类的准确率(见表 3)。

4.3. 实验的计算成本

本文中的所有仿真实验用 Matlab 2012a 中实现, 在个人电脑的配置为英特尔(R)酷睿(TM)I5-2400, CPU 的频率 3.10 GHz 的, 运行内存 3.00 GB, 操作系统是微软 Windows XP 专业版。我们在本实验中, 只计算一下各个算法的运行时间。从时间复杂度的角度证明算法。对于每一种算法, 我们严格的根据所提供的论文中的方法进行实验。在我们的方法中, 极光图像转换成相应的灰度级图像, 极大的降低了计算的复杂度, 但并不是只针对我们的算法进行改变, 所有的分类算法都进行了灰度级的转换。

实验三: 比较各种方法的计算成本(s) (见表 4):

在极光的分类算法中, 有几种已经取得了不错的分类效率。本节的实验目的就是用比较本文提出的算法和先前的几类算法进行三方面的比较, 在算法效率的前提下对另外的一个非常重要的问题即抗噪性能方面做比较。在很多研究领域, 分类算法在不加噪声的情况下, 分类的效率可以达到非常的高, 但是现实中, 任何情况下都会有噪声的干扰。其中, 高斯噪声和椒盐噪声是在实际的问题中碰到的最多的噪声类型, 用这两种噪声进行干扰的目的是为了适应更复杂的实际情况。因此, 抗噪的性能已经越来越成

Table 1. Classification efficiency of various classification algorithms (%)
表 1. 各种分类算法的分类效率(%)

	arc	hotspot	冕状	小尺度
基于极光弧的极光识别分类方法	83.05	90.05	77.35	81.54
K-阶近邻方法	77.94	90.98	86.38	82.85
BP 神经网络	79.92	88.38	85.25	80.07
LBP-SVM	93.84	86.52	77.13	91.42
本文的方法	94.40	95.55	92.02	81.33

Table 2. Various classification algorithms for Gaussian noise robustness (%)
表 2. 各种分类算法对高斯噪声的鲁棒性(%)

	arc	hotspot	冕状	小尺度
基于极光弧的极光识别分类方法	79.52	80.06	73.65	70.20
K-阶近邻方法	75.22	85.03	81.66	77.35
BP 神经网络	78.22	81.33	82.86	74.95
LBP-SVM	88.96	84.21	73.33	90.12
本文的方法	92.22	88.16	91.78	90.73

Table 3. Various classification algorithms for impulse noise robustness (%)
表 3. 各种分类算法对椒盐噪声的鲁棒性(%)

	arc	hotspot	冕状	小尺度
基于极光弧的极光识别分类方法	77.46	75.33	70.38	66.54
K-阶近邻方法	74.32	85.01	79.23	70.22
BP 神经网络	73.36	80.05	71.58	70.22
LBP-SVM	84.23	83.66	71.23	86.39
本文的方法	91.55	84.11	88.74	89.22

Table 4. Algorithm's computational cost (s)
表 4. 算法的计算成本(s)

实验方法	运行时间
基于极光弧的极光识别分类方法	1.1273
K-阶近邻方法	0.6836
BP 神经网络	2.7029
LBP-SVM	1.2134
本文的方法	1.8163

为验证一种新的分类算法的性能指标。最后，算法的实际的计算成本，我们本实验只从算法的实际运行的时间角度对算法的成本做一种简单的研究。以此来模拟实际的计算成本。

实验的结论：

1) 在分类效率的部分，我们通过对最初的极光识别分类算法和现有的比较主流的分类识别算法的分

类效率进行比较, 从实验的角度论证了本文所提出的分类效果比目前提出的几种分类的效率要好一点。

2) 在预处理后的极光图像后加入高斯噪声(均值 0, 方差 0.01)和椒盐噪声(密度 0.05), 发现本文的方法相对于其他的几种方法的抗噪性能好, 而且也很稳定。

3) 通过最后的实验也表明, 本文的所提出的方法的实验成本不是最好的, 但是也在能接受的范围之内。

5. 结论

由于极光数据的庞大和变化的频率非常的快, 极光数据的变化规律存在着大量的不确定性。传统法的模式识别技术不能在分类效率和抗噪声性能方面极大的提升。要想达到预期效果, 必须经过提出或者改进新的模式识别的技术。本文通过提出自适应的小波提升方案和改进的局部二值模式, 最后用模糊的邻近分类相结合的方法用于极光图像的分类试验。通过仿真的实验, 我们的算法效率在分类效率鲁棒性的方面, 都比之前的算法有了提高。另外在实验成本方面, 本文的算法也在可以接受的范围之内。在下一步工作中首先, 将会继续突破自适应提升小波变换的三个方向的限制, 使得从各个方向都能进行选取, 其次就是会引入一些非监督的分类方法, 进一步的提高分类的效率。

参考文献 (References)

- [1] Wang, Q., Liang, J., Hu, Z.J., Hu, H.H., Zhao, H., Hu, H.Q., Gao, X.B. and Yang, H.G. (2010) Spatial Texture Based Automatic Classification of Dayside Aurora in All-Sky Images. *Journal of Atmospheric and Solar-Terrestrial Physics*, **72**, 498-508. <http://dx.doi.org/10.1016/j.jastp.2010.01.011>
- [2] Afshang, M., Helfroush, M.S. and Zahernia, A. (2009) Gabor Filter Parameters Optimization for Texture Classification Based on Genetic Algorithm. *Proceedings of the 2nd International Conference on Machine Vision*, Dubai, 28-30 December 2009, 199-203. <http://dx.doi.org/10.1109/icmv.2009.50>
- [3] 王倩, 梁继民, 高新波, 等. 基于表象特征的极光图形分类方法研究[C]//中国空间科学学会. 第十二届全国日地空间物理学术研讨会论文集. 2007, 7, 1.
- [4] 高凌君, 高新波, 梁继民, 等. 一种基于 Gabor 变换的日侧静态极光分类算法[C]. 第七届全国信号与信息处理联合会议暨首届全国省(市)级图象图形学会联合年会. 2008.
- [5] Fu, R., Gao, X., Li, X., et al. (2010) An Integrated Aurora Image Retrieval System: Aurora Eye. *Journal of Visual Communication and Image Representation*, **21**, 787-797. <http://dx.doi.org/10.1016/j.jvcir.2010.06.002>
- [6] Wang, X., Liang, J.H. and Guo, F.X. (2014) Feature Extraction Algorithm Based on Dual-Scale Decomposition and Local Binary Descriptors for Plant Leaf Recognition. *Digital Signal Processing*, **34**, 101-107. <http://dx.doi.org/10.1016/j.dsp.2014.08.005>