Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Matching methods evaluation framework for stereoscopic breast x-ray images

Johanna Rousson
Mathieu Naudin
Cédric Marchessoux

Matching methods evaluation framework for stereoscopic breast x-ray images

Johanna Rousson,* Mathieu Naudin, and Cédric Marchessoux

Barco NV, Healthcare Division, President Kennedypark 35, Kortrijk 8500, Belgium

Abstract. Three-dimensional (3-D) imaging has been intensively studied in the past few decades. Depth information is an important added value of 3-D systems over two-dimensional systems. Special focuses were devoted to the development of stereo matching methods for the generation of disparity maps (i.e., depth information within a 3-D scene). Dedicated frameworks were designed to evaluate and rank the performance of different stereo matching methods but never considering x-ray medical images. Yet, 3-D x-ray acquisition systems and 3-D medical displays have already been introduced into the diagnostic market. To access the depth information within x-ray stereoscopic images, computing accurate disparity maps is essential. We aimed at developing a framework dedicated to x-ray stereoscopic breast images used to evaluate and rank several stereo matching methods. A multiresolution pyramid optimization approach was integrated to the framework to increase the accuracy and the efficiency of the stereo matching techniques. Finally, a metric was designed to score the results of the stereo matching compared with the ground truth. Eight methods were evaluated and four of them [locally scaled sum of absolute differences (LSAD), zero mean sum of absolute differences, zero mean sum of squared differences, and locally scaled mean sum of squared differences] appeared to perform equally good with an average error score of 0.04 (0 is the perfect matching). LSAD was selected for generating the disparity maps. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.3.1.011007](https://doi.org/10.1117/1.JMI.3.1.011007)]

Keywords: stereoscopy; displays; medical imaging; image quality; three dimensions; computer vision.

Paper 15133SSR received Jul. 2, 2015; accepted for publication Oct. 6, 2015; published online Nov. 18, 2015.

1 Introduction

1.1 Three-Dimensional in Breast Screening

The benefit of using three-dimensional (3-D) over two-dimensional (2-D) images is the added depth information. With stereomammography, the depth information is provided by stereoscopic viewing of two 2-D mammograms, acquired at two different angles (typically separated by 4 deg), on a 3-D display such that each eye receives only one of the two mammograms. The fusion into the brain of these two slightly angularly separated mammograms (i.e., stereomammograms) generates a 3-D representation of the breast.

Depth information from stereoscopic viewing should offer to radiologists a more realistic spatial orientation and a better understanding of tissue distributions and volume shapes. In fact, 3-D x-ray imaging could improve breast cancer diagnosis. Some previous work has shown a potential clinical benefit of a 3-D full imaging system (i.e., stereomammography) over with typical digital 2-D x-ray systems.¹ In 2008, Getty et al.¹ compared a 3-D full imaging system consisting of a stereoscopic x-ray imaging system with a passive stereoscopic medical display system and a typical digital 2-D x-ray system consisting of a digital x-ray detector and a medical display system. A total of 1458 cases were used. The 3-D full imaging system reduced false positive lesion detection by 46% ($p < 0.0001$), and significantly increased true positive lesion detection by 23%

($p < 0.05$). A similar study handled by the same group was published in 2013 with 1298 cases.² The 3-D system significantly improved false positive fraction with a specificity of 91.2% for the 3-D system and 87.8% rate for the 2-D system. “We found that the stereoscopic technique could significantly decrease the need for calling women back for additional exams,” Dr. D’Orsi said. However, unlike surgeons, radiologists are still not familiar with 3-D display imaging systems for diagnostic purposes and are not convinced by the added or clinical value of the depth information. By instance, for breast cancer screening, the current standard is 2-D digital mammography^{3,4} even if 3-D x-ray acquisition systems and 3-D medical displays have already been introduced into the market.^{3,5}

Additionally, using stereomammography may reduce the needs of multimodalities and thus the amount of the total radiation dose. Currently, there are a lot of concerns about the reduction of radiation dose in radiography. New breast modalities like digital breast tomosynthesis (DBT)⁶ may not be sufficient to detect some types of lesions such as microcalcifications. Consequently, the visualization of 2-D digital mammography images on 2-D medical displays may still be required in addition to these new modalities,⁷ increasing the total radiation dose. Stereomammography has been proven to circumvent drawbacks of 2-D digital mammography (e.g., overlapping of structures) by providing depth information (like DBT), without increasing the radiation dose (i.e., 1 to 1.1 times the dose required for one 2-D digital mammogram).^{8,9}

*Address all correspondence to: Johanna Rousson, E-mail: johanna.rousson@barco.com

1.2 Stereo Matching Evaluation

In past decades, computer-aided detection (CAD) systems dedicated to breast screening have been developed to assist radiologists in making more accurate diagnosis. CAD tools have been demonstrated to increase the detection of true positive breast cancers.^{7,10–12} Typically, CAD systems (1) process the mammograms to extract features from regions of interest (e.g., suspicious areas) and (2) display markers to highlight these regions. Current CAD systems are not suitable with stereomammography since they are not designed to compute and account for depth information.

This depth information can be retrieved by knowing both the disparity (i.e., the difference in location) between corresponding points in the two stereoscopic images and the acquisition geometry. A dense disparity map is a 2-D image containing the different disparities between all the stereo matched points of the stereoscopic images. If the imaging acquisition system has a particular epipolar geometry, called coplanar, where both left and right matching pixels are on the same row or line in each image of the stereo pair, then only one vector information (with its origin point and its norm) is needed for generating the disparity map. The vector corresponds to the vector between the matched points in the left and in the right images. Consequently, to produce the disparity map, the norm of this vector, combined to its direction (i.e., either a positive or a negative sign in front of the norm since the images are epipolar), should be saved at the pixel position corresponding to the origin position of the vector. Otherwise, either both the origin point and the end point of the vector or the origin point, the norm, and the angle (or the direction) should be saved.

To compute such disparity maps, stereo matching techniques have to be developed for matching the points of the stereoscopic images. Despite stereo matching techniques having been studied for many years in photography,¹³ limited efforts have been devoted to the stereo matching of stereoscopic medical,¹⁴ and especially breast, x-ray images.^{7,15} This is partly due to the complexity of x-ray breast images, to the limited number of available stereomammography systems in the world, and therefore to the limited access to real stereomammograms by research groups. Indeed, to the authors knowledge, most research groups have been working on either phantom-based images or images acquired with tomosynthesis.^{7,15} Therefore, more studies have to be done to develop stereo matching algorithms regarding stereomammograms.

Furthermore, it would be very valuable to not only develop stereo matching algorithms for stereomammography, but also to develop stereo matching algorithms applicable to both 3-D modalities (i.e., tomosynthesis and stereomammography) to generate interoperable CAD systems.

To assess the efficiency of stereo matching methods, evaluation frameworks can be used. Several evaluation frameworks for comparing stereo matching methods exist. A well-known one is the Middlebury framework¹³ including a large dataset of stereo pairs with their ground truth disparity maps. Several stereo matching and disparity map generation methods are available. The complete framework can be downloaded from the Internet.¹⁶ Unfortunately, in all cases, the images are photography images and the ground truth of the disparity maps is a simple collection of segmented planes, meaning that the 3-D scene simply includes a set of different planes, i.e., is made of piecewise smooth areas. Other frameworks do exist but have similar limitations. For breast images, having only a set of planes is out

of scope. Therefore, a dedicated framework must be developed for x-ray breast images.

In this article, we aim at generating a framework used to assess and rank stereo matching methods, when both phantom-based tomosynthesis breast x-ray images and stereomammograms are considered. Our final goal is to (1) extract the depth information from stereoscopic images and (2) to develop new visualization techniques, e.g., to measure the depth and the radius of a tumor within the breast, to detect microcalcifications and to develop CAD systems. Therefore, the stereo matching framework is also designed to generate a dense disparity map from a given stereo matching method.

Section 2 describes the specificities of x-ray breast images and why a dedicated framework has to be developed. The stereo matching methods included in the framework are defined in Sec. 3. Section 4 presents the definition and the development of an automatic C++ evaluation framework with different parts. The first part deals with the description of the dataset that has been selected and used for the evaluation. The second part focuses on the explanation of the ground truth that we elaborated for evaluating the performance of the framework including a dedicated tool for generating the ground truth. Finally, the optimization and the implementation of the selected matching methods, their tuning parameters, and finally the metric definition for evaluating the performance of the stereo matching techniques are described. Results of the evaluation are given in Sec. 5 and discussed in Sec. 6.

2 Challenges with X-Ray Breast Images

A mammogram is a simple breast radiography for which the breast has been compressed between two plates (Fig. 1). Therefore, the thickness or the depth of the breast is limited (a few centimeters). It is very important to compress the breast to spread the tissue apart and therefore to allow the maximum amount of tissue to be imaged and to reduce radiation dose. The breast size and compressed volume are important parameters as

Fig. 1 Illustration of the geometry of the acquisition system. The two stereo images are obtained by performing acquisition 1 and acquisition 2, respectively. Angle α is between the two acquisitions 1 and 2. Thickness corresponds to the distance between the detector and the plate used to hold the three-dimensional (3-D) object under measurement (e.g., the breast), i.e., the thickness of the measured 3-D object. MaxDisp is the largest distance in pixels between two corresponding points in the stereo pair.

well as the breast density. The compression of the breast limits the depth of the breast and therefore the 3-D information. In stereomammography, two 2-D mammograms are acquired at two different angles (typically separated by 4 deg), as illustrated in Fig. 1.

The intensity values in a 2-D x-ray image are proportional to the x-rays transmitted to the object which is being imaged. Changing the image acquisition angle results in a different projection, due to the difference in the object interior along the x-ray path. The situation is different for photography, in which changing the acquisition angle only results in a different region at the object surface which is seen by the camera. In fact, in radiography, images taken at different angles may differ even when the regions of the object surface stay the same, since the projections are generated by the superposition of the x-ray attenuation through different portions of the object interior. These differences between the projections increase with the increasing angle between two acquisitions. Therefore, if this angle remains small, the differences are likely to be limited (but still exist), and some of the state-of-the-art stereo matching methods designed for photography (e.g., local methods, see Sec. 3) may also work for radiography.

Finally, x-ray breast images characterized by tissue density are made of a combination of different structures (e.g., fibroglandular tissue, fat tissue) and can vary significantly between individuals resulting in a challenge for matching points between pairs of such images. These are big challenges that the x-ray stereo matching techniques must handle.

Another challenge is the resolution in terms of pixels. An x-ray detector can include more than 10 million pixels making the implementation and the time to process such high-resolution images challenging.

3 Assessed Stereo Matching Methods

Local, global, and hybrid stereo matching methods have already been studied for many years. Local methods are used for dense disparity extraction. A window (also named kernel) is calculated from the left image where the centered pixel is the pixel to be matched with a pixel of the right image. A window in the right image is also calculated and moved over the image to compute the matching cost with the left window. Matching costs or correlation scores are used to estimate the best matching window between the left and the right images. These techniques use pixel intensity as attribute. For cost calculation, the smallest cost means the best candidate. For correlation, the highest score means the best candidate.

The following local stereo matching methods have been selected for evaluation with our framework since they are the state-of-the-art of stereo matching techniques for photography and may work for matching stereomammograms. These techniques have been used successfully for different applications in photography such as in industrial context. They are all based on convolution and result in a disparity map made of a disparity value $\text{disp}(x, y)$ as pixel information. With $\text{disp}(x, y)$ defined as

$$\text{disp}(x, y) = \text{argmax}[C_{x',y'}(\text{disp})], \quad (1)$$

where C is the correlation function assigning to each pixel position a correlation score between the pixel (x, y) in the left stereo image and each pixel (x', y') within the search window disp in the right stereo image. The maximal value means the position of the best candidate within the evaluated window is returned, and

the difference in location between the left and the right stereo matched points is provided to the pixel (x, y) of the disparity map. For cost calculation, the function C assigns a cost to each pixel position and therefore each disparity value $\text{disp}(x, y)$ is defined as

$$\text{disp}(x, y) = \text{argmin}[C_{x',y'}(\text{disp})]. \quad (2)$$

The local stereo matching methods included in the framework are:

1. Metrics based on absolute differences:

- Sum of absolute differences (SAD)^{17–19}
- Zero mean sum of absolute differences (ZSAD)^{18,19}
- Locally scaled sum of absolute differences (LSAD)^{19,20}

2. Metrics based on squared differences:

- Sum of squared differences (SSD)^{18,19,21}
- Zero mean sum of squared differences (ZSSD)^{18,19}
- Locally scaled mean sum of squared differences (LSSD)¹⁹

3. Metrics based on cross correlation:

- Normalized cross correlation (NCC)^{18–20,22}
- Zero mean normalized cross correlation (ZNCC)^{18,19}

Additionally, several global methods exist. They use several mathematical models or theories such as Markov random field or graph theory. In order to decrease the searching difficulty, global methods only use a one-dimensional search. In order to test one of the global methods, the loopy belief propagation (LBP)²³ has been selected, but knowing upfront that this technique may not fit the needs and constraints with breast x-ray images that were explained previously. There are also many hybrid methods^{24,25} that use other information than the intensity of the pixels or are based on Euler–Lagrange theory²⁶ by instance. These methods were not selected due to the aforementioned needs and constraints of breast x-ray stereoscopic images.

4 Evaluation Framework

4.1 Dataset

In order to test and select the most robust stereo matching method, a variety of breast datasets have been selected.

4.1.1 Real dataset

In total, 80 real 2-D stereoscopic x-ray mediolateral oblique (MLO) and craniocaudal (CC) images from 20 patients are available. From which, 40 pairs of stereoscopic breast images can be used.

The stereomammography acquisition system used in this study represents the state-of-the-art,⁵ i.e., the two mammograms are acquired at two different angles separated by 4 deg. Additionally, to not have the total dose exceeding the maximum authorized radiation dose, one of the two mammograms is acquired with less dose, and therefore comprises more noise

Fig. 2 Cropped real stereomammograms: (a) left, and (b) right. The right stereomammogram is acquired with less radiation dose than the left stereomammogram. The left and right mammograms were acquired with a separation angle of 4 deg.

than the other mammogram (see Fig. 2). Acquiring the two stereomammograms with two different radiation doses is viable since the human brain can properly reconstruct the 3-D, although the stereo images comprise different levels of noise.⁹

4.1.2 Semivirtual dataset

The second set is a semivirtual set, i.e., a breast phantom (not simulated) was used in combination with a real image acquisition system. The acquisition system is the same as for the real dataset with a separation angle of 4 deg between the two stereomammograms. For the breast under study, the phantom is not real, so two full radiation doses are applied. Images are less noisy and therefore have a better image quality. Four 2-D stereoscopic x-ray MLO and CC images acquired from the same phantom are available, from which two pairs of stereoscopic images can be extracted. Cropped versions of the images of the semivirtual dataset are illustrated in Fig. 3.

4.1.3 Virtual dataset

The virtual dataset was generated using the virtual clinical trials platform developed at the University of Pennsylvania.^{27,28} Both the breast, or the phantom, and the image acquisition system are simulated. The dataset is a simulation of DBT based on a virtual anthropomorphic phantom. The images correspond to the DBT's projections. The angle between two consecutive projections is equal to 2.67 deg. From the set of DBT's projections, several stereoscopic pairs can be generated (see Fig. 4). Twenty-six projections from two simulations are available. Three projections emphasizing the generation of disparity with the camera movement are given in Fig. 5. The two sets comprise 15 projections. From these two sets, several combinations of stereo pairs with different angles of separation, i.e., with an angle of 2.67 deg or a multiple of 2.67 deg, can be generated.

The stereo matching methods will be evaluated with the quality of correct matching of pixels or points between the left and

Fig. 3 Cropped semivirtual stereomammograms: (a) left, and (b) right. The right stereomammogram is acquired with the same radiation dose as the left stereomammogram. The left and right mammograms were acquired with a separation angle of 4 deg.

right images. Therefore, in order to evaluate the correct matches, a ground truth with known matching points is required.

4.2 Ground Truth

The ground truth corresponds to the matching points between the reference image (e.g., the left image) and the compared one (e.g., the right image). There are several ways for generating a ground truth: fully automatic, semiautomatic, or manual. A dedicated interface has been developed in Qt for opening, zooming, and panning a pair of images and also generating the ground truth in two modes: manual and semiautomatic.

Fig. 4 Cropped virtual stereo digital breast tomosynthesis (DBT) projections: (a) left, and (b) right.^{27,28} The left and right projections were acquired with a separation angle of 2.67 deg.

Fig. 5 Three projections of a simulated DBT, projection 00, 07, and 14.^{27,28} The projections correspond to the x-ray tube positions during the acquisition of tomosynthesis projections: (a) leftmost, (b) central, and (c) rightmost. To highlight the camera movement versus the generated disparity, a point present in the three projections is marked with a cross.

In the manual mode, called “sharp points,” the user selects points, for which he or she easily finds a correspondence more or less in the other image. A corresponding cross with a number is displayed on both images. The user can move the cross and place it at the correct position.

In the semiautomatic mode, called “linear grid,” a linear grid of crosses is drawn automatically with a given step size on the reference image and on the compared one. The user can move the points in the same way as in the manual mode for generating the ground truth.

Attention has been made in order to facilitate the use of the software and to bring about a nice user experience.

Figures 6 and 7 show screenshots of the ground truth interface. Figure 6(a) is a screenshot of the interface when a pair of images is opened. Figure 6(b) is a screenshot of the interface in manual mode. Figure 7(a) is a screenshot of the interface in semiautomatic mode. Figure 7(b) is a screenshot of the interface zoomed in. The green crosses correspond to the reference points (e.g., in the left image) and the red crosses correspond to the points in the other image (e.g., the right image).

A dedicated binary format has been defined for saving the ground truth with a dedicated data structure including the image width and height, number of matched points, and vectorial information between matched points.

Fig. 6 Several functionalities of the ground truth interface: (a) files opening, and (b) manual mode with sharp points; images obtained with the breast phantom simulation from the University of Pennsylvania.^{27,28} The green crosses correspond to the reference points (e.g., in the left image) and the red crosses correspond to the points in the other image (e.g., the right image).

Fig. 7 Several functionalities of the ground truth interface: (a) semiautomatic mode with a linear grid of points and (b) zoomed in; images obtained with the breast phantom simulation from the University of Pennsylvania.^{27,28} The green crosses correspond to the reference points (e.g., in the left image) and the red crosses correspond to the points in the other image (e.g., the right image).

4.3 Stereo Matching C++ Framework and Its Optimization

The stereo matching framework, illustrated in Fig. 8, was developed to implement stereo matching methods and produces a result being either a file comprising the matching results for a limited number of points or a disparity map. The main purpose of the framework is to find for each pixel of the reference (left) image a corresponding pixel in the test (right) image. Occlusions are handled by the framework. The dense stereo matching framework is the part of the main framework used to compute a dense disparity map. In a dense disparity map, all pixels from the image are evaluated (dense disparity map extracted). The sparse stereo matching framework is used to compute and to evaluate sparse stereo matched points. In sparse stereo matching, only the points with the same coordinates as the points of the ground truth file are evaluated.

4.3.1 Design

“Design pattern” methodology is a prerequisite for any oriented programming languages. In combination with classical unified modeling language representation, design pattern is used for establishing an optimized C++ design answering to the constraints and the needs of this framework. The images are large and convolution is necessary for most of the stereo matching methods. A factory (MethodsFactory) has been used to automatically substantiate the stereo matching methods to test. A dedicated data structure has been developed for organizing and saving the ground truth data.

4.3.2 Multiresolution optimization

The multiscale image optimization, or pyramid, is used to be more accurate and more efficient in our matching search. In

Fig. 8 Block diagram of the stereo matching framework.

Fig. 9 Multiscale pyramid: (a) multiscale representation (3 levels), and (b) example of a point found in a two-level pyramid.

order to create the image pyramid, each level size is divided by 2 [cf. Fig. 9(a)]. To create the image, pixels are interpolated by using triangular interpolation because it is a very fast and efficient method [cf. Fig. 9(b)]. This optimization helps the algorithm by limiting the research zone.

The maximum disparity is the largest possible distance in pixels between two corresponding points in the stereo pair. The maximum disparity [MaxDisp in Eq. (3) and Fig. 1] is computed accounting for the geometry of the acquisition system, i.e., the angle of acquisition [α in Eq. (3) and Fig. 1] between the two stereo images, the size of a pixel of the detector [PixSize in Eq. (3)], and the thickness of the object under study [Thickness in Eq. (3) and Fig. 1].

The maximum disparity is recalculated for the top level in order to search in the desired area. To be as accurate as possible, the window size is increased by two pixels in all directions. This window size increase is a margin precaution called tolerance zone. The process for finding a point from the top pyramid level (low-resolution image) to the bottom level (high-resolution image) is not trivial. The position of the reference point is found at low resolution [with Eq. (4)] and defines the center of the searching zone in the right position in the pyramid. The algorithm will find a point [Fig. 9(b)] and go to the next level (Current level - 1). Due to the interpolation, several pixels are candidates in the new level (at least four). The tolerance zone is drawn in yellow in Fig. 9(b). The new searching block includes both the four candidates and the tolerance zone. Then, in this new zone the algorithm will proceed the same as for the first level, and will return a point. Therefore, this process will return a point at high resolution correlated to the preceding found points. If a bad match is made on the top level, the error will be reflected on each level. The tolerance zone was created to avoid that the correct target point is excluded from the searching zone.

$$\text{MaxDisp} = \frac{2 \text{ Thickness } \tan\left(\frac{\alpha}{2}\right)}{\text{PixSize}}, \quad (3)$$

$$\text{Position}_{\text{level}N} = \text{HighResolutionCoordinates} (0.5)^N. \quad (4)$$

If several candidates are found for the same reference point, then none of them is selected and this results in an occlusion point.

4.3.3 Input and output

At the first input the stereo matching evaluation framework (Fig. 8) needs an .ini file with a list of the methods to test with their parameter values such as the kernel size, the maximum disparity, and the number of pyramid levels. The second input is also an .ini file containing the full path of the reference image, the compared image, and the ground truth file. Each line represents a stereoscopic pair to process. The matching points are then saved in the same format as the ground truth format described in Sec. 4.2. The evaluation can be executed for comparing the ground truth to the outcome of a stereo matching technique based on a metric.

4.3.4 Metric

To compare the stereo matched point (p_{Res} in Fig. 10) returned by the different matching methods to the ground truth stereo matched point (p_{GT} in Fig. 10) a dedicated metric was designed. This metric, depicted in Eq. (8), aims at scoring the distance between p_{Res} and p_{GT} . The larger the score, the larger the distance between p_{GT} and p_{Result} and the larger the difference. The reference point in the left image is defined as p_{Ref} .

The metric comprises two components:

- A score [$\text{Norm}_{\text{Error}}$ in Fig. 10(a) and Eq. (5)] evaluating the Euclidean distance between p_{Res} and p_{GT} [$\text{norm}_{\text{vGTRes}}$ in Eq. (5)] normalized by the maximum reachable distance [$\text{norm}_{\text{vGTErrMax}}$ in Eq. (5)], i.e., the Euclidean distance between p_{GT} and p_{ErrMax} .

Fig. 10 Metric calculation parameters for computing: (a) $\text{Norm}_{\text{Error}}$, and (b) $\text{Angle}_{\text{Error}}$; Angle corresponds to the angle $pGTpRefpRes$.

$pErrMax$ is the point the furthest from pGT in the searching stereo matching point area, i.e., the zone centered on the reference point ($pRef$) and delimited by the maximum disparity (Fig. 10).

- A score [$\text{Angle}_{\text{Error}}$ in Eq. (7)] assessing the angular deviation [angle in Fig. 10(b) and Eqs. (6) and (7)] between the two vectors $\vec{vRefRes}$ and $\vec{vRefGT}$ [Eqs. (6) and (7)] normalized by the maximum deviation, i.e., 180. $\text{Angle}_{\text{Error}}$ is used to evaluate whether $pRes$ is found along a direction close to the direction of pGT , i.e., whether $pRes$ follows the direction of the movement of the x-ray system. An angle of 180 deg means that $pRes$ is found along the opposite direction of pGT (which corresponds to the worst stereo matching case).

$$\text{Norm}_{\text{Error}} = \frac{\text{norm} \vec{vGTRes}}{\text{norm} \vec{vGTErrMax}}, \quad (5)$$

$$\vec{vRefRes} \cdot \vec{vRefGT} = \|\vec{vRefGT}\| \cdot \|\vec{vRefRes}\| \cdot \cos(\text{Angle}), \quad (6)$$

$$\text{Angle}_{\text{Error}} = \frac{\text{Angle}}{180}. \quad (7)$$

Finally, the metric is the weighted sum of $\text{Norm}_{\text{Error}}$ and $\text{Angle}_{\text{Error}}$ as described in Eq. (8) with w_n and w_a the weights applied to $\text{Norm}_{\text{Error}}$ and $\text{Angle}_{\text{Error}}$, respectively.

$$\text{Weighted}_{\text{Error}} = \frac{w_a \cdot \text{Angle}_{\text{Error}} + w_n \cdot \text{Norm}_{\text{Error}}}{w_a + w_n}. \quad (8)$$

To properly select w_n and w_a , several stereo matched points were compared to a single ground truth point. Twelve right points have been defined at several pixel distances and several angle values from the correct right point in order to evaluate which weights gave the best results of the metric. These twelve right points have been compared to the ground truth point

regarding four different left points covering four quadrants of 90 deg in order to reflect which attribute is the most important: the angle or the distance. All these point combinations have been tested with different combination values of w_n and w_a (w_n from 0 to 1 and w_a from 1 to 0). As the evaluated right points were known, the ranking from the best match to the worst match was known as well. The norm weight should be more important than the angle weight. Nevertheless, the angle weight should still penalize points that are not in the direction of the movement of the x-ray system from the left image to the right image. The weights combination returning the best ranking compared to the known ranking was $w_n = 0.7$ and $w_a = 0.3$.

The metric was mainly used to rate stereo matched points at the level 0 of the pyramid.

5 Results

5.1 Local Stereo Matching Methods Comparison

To compare the eight local stereo matching methods enumerated in Sec. 1, the error score returned by the metric is computed for all the individual stereo matched points at level 0 (i.e., high resolution) of the same input image type. For all assessed stereo matching methods, the standard deviation, as well as the mean, the maximum, and the minimum of the error scores (computed over 130 points on average per image type) are calculated per image type (i.e., real, semivirtual, and virtual image types). Additionally, a principal component analysis (PCA)²⁹ is conducted to sort out the different methods. The standard deviation, as well as the mean, the maximum, and the minimum of all the error scores computed over the three aforementioned image types (depicted in Table 1) are given as variables to the PCA. The GroundTruth point was added to the PCA as a method whose variables were all set to 0.

Figure 11 is a map representing the directions along which the different methods vary as well as their spread for the two first components. Results of the PCA suggest that ZNCC and NSSD methods (numbers 8 and 9, respectively, in both Table 1 and Fig. 11; the number 1 corresponds to the GroundTruth) are two outliers and perform worse than the other stereo matching methods, with error scores of 0.149 and 0.345 on average. All the other methods appear to perform equally good although

Table 1 Mean, standard deviation, minimum, and maximum of all the error scores computed over the three different image types at level 0 of the pyramid (i.e., high resolution). The mean, standard deviation, minimum, and maximum are calculated for eight different local stereo matching methods and two parameters, i.e., a 13×13 convolution kernel and a maximum disparity of 60.

Level 0					
n#	Methods	Mean	Std	Min	Max
1	GT	0	0	0	0
2	SAD	0.0547	0.0527	0.0013	0.271
3	SSD	0.0523	0.0517	0.0006	0.267
4	ZSAD	0.0401	0.0428	0.0003	0.2176
5	ZSSD	0.0398	0.0398	0	0.2444
6	LSAD	0.0377	0.038	0	0.1914
7	LSSD	0.0393	0.0405	0	0.2492
8	ZNCC	0.1487	0.1171	0.0074	0.4753
9	NSSD	0.3447	0.1029	0.0806	0.5473

Note: Bold value denotes the result of the method with the best performance (per column).

Fig. 11 Plot of the spread of the evaluated stereo matching methods for the first two components of the principal component analysis. Numbers represent the different stereo matching methods as given in Table 1. The number 1 corresponds to the GroundTruth, i.e., an average mean, average standard deviation, average minimum, and average maximum error score set to 0.

the LSAD method (number 6 in both Table 1 and Fig. 11) is the closest to the GroundTruth point (number 1 in Fig. 11) with an error score of 0.038 on average.

A PCA is also carried out for each image type to compare stereo matching performance between the three different

Fig. 12 Plot of the spread of the evaluated stereo matching methods for real image type for the first two components of the principal component analysis. Numbers represent the different stereo matching methods as given in Table 1. The number 1 corresponds to the GroundTruth, i.e., a mean, a standard deviation, a minimum, and a maximum error score set to 0.

Fig. 13 Plot of the spread of the evaluated stereo matching methods for semivirtual image type for the first two components of the principal component analysis. Numbers represent the different stereo matching methods as given in Table 1. The number 1 corresponds to the GroundTruth, i.e., a mean, a standard deviation, a minimum, and a maximum error score set to 0.

datasets. As depicted in Figs. 12 through 14 for the real, semi-virtual, and virtual datasets, respectively, ZNCC and NSSD methods perform the worst, and thus should not be selected to search for stereo matching points in stereoscopic breast

Fig. 14 Plot of the spread of the evaluated stereo matching methods for virtual image type for the first two components of the principal component analysis. Numbers represent the different stereo matching methods as given in Table 1. The number 1 corresponds to the GroundTruth, i.e., a mean, a standard deviation, a minimum, and a maximum error score set to 0.

x-ray images. LSAD, ZSAD, LSSD, and ZSSD methods all appear to be good candidates for all three image types with the two tested angles of acquisition (i.e., 2.67 deg and 4 deg), as illustrated in Table 2. Nevertheless, scores given in Table 2 suggest that LSSD, ZSSD, and LSAD are the most appropriate stereo matching methods regarding the real, semivirtual, and virtual datasets, respectively. Finally, the four stereo matching techniques considered in Table 2 are more efficient regarding the semivirtual dataset than the real dataset. Since they were acquired with two full radiation doses, the semivirtual stereomammograms (Fig. 3) are less noisy than the real stereomammograms (Fig. 2), hence the increase of the stereo matching efficiency.

Since LSAD is the stereo matching method, the closest (i.e., with the smallest error) to the GroundTruth for the three datasets taken together (Fig. 11), we generated several disparity maps for the three different image types using the LSAD method (with a 13×13 convolution kernel and a pyramid with seven levels). Occlusions are reference points for which not a single candidate has been found. The number of occlusions was always less than 0.1% of the total number of pixels. This means that for one reference point, the stereo matching method almost always finds one single candidate (note that the correctness of the stereo matching is not accounted for). When occlusions are found, the returned disparity is set to a default value and thus should be considered as a hole to be filled by some postprocessing.

An example of a disparity map generated with the LSAD method applied on left and right stereo x-ray breast images extracted from the virtual dataset is given in Fig. 15. The disparity map was produced using a convolution kernel of 13×13 pixels and seven pyramid levels.

Table 2 Mean, standard deviation, minimum, and maximum of all the error scores computed for the three different datasets at level 0 of the pyramid (i.e., high resolution). The mean, standard deviation, minimum, and maximum are calculated for zero mean sum of absolute differences (ZSAD), zero mean sum of squared differences (ZSSD), locally scaled sum of absolute differences (LSAD), locally scaled mean sum of squared differences (LSSD; i.e., the best four), local stereo matching methods, and two parameters, i.e., a 13×13 convolution kernel and a maximum disparity of 60.

Level 0						
Datasets	n#	Methods	Mean	Std	Min	Max
Real	4	ZSAD	0.0401	0.0518	0	0.2898
	5	ZSSD	0.035	0.0382	0	0.2110
	6	LSAD	0.0348	0.04	0	0.2112
	7	LSSD	0.0332	0.0381	0	0.2110
Semivirtual	4	ZSAD	0.0236	0.0246	0	0.1968
	5	ZSSD	0.0211	0.0197	0	0.1825
	6	LSAD	0.0222	0.0218	0	0.1968
	7	LSSD	0.0212	0.0223	0	0.1968
Virtual	4	ZSAD	0.0565	0.052	0.0009	0.1662
	5	ZSSD	0.0634	0.0614	0	0.3398
	6	LSAD	0.0563	0.0522	0	0.1662
	7	LSSD	0.0636	0.0612	0	0.3398

Note: Bold value denotes the result of the method with the best performance (per column).

5.2 Multiresolution Optimization

As described in Sec. 4.3.2, the multiscale image optimization is used to have a matching search more accurate and more efficient. Indeed, instead of searching in a large square zone whose width is the max disparity (i.e., 120 pixels), the multiresolution optimization focuses the search in a much limited zone (19×19 pixels at the lowest resolution and 6×6 pixels at the other levels). Decreasing the size of the searching zone increases the efficiency by at least a factor eight, but also reduces the number of candidates and thus increases the accuracy. As depicted in Table 3, for each image type (real, semivirtual, and virtual) the mean error score has been improved by about 70% for the virtual images and by more than 90% for the real and semivirtual images by using the multiscale image optimization compared with the results of the stereo matching computation performed without the aforementioned optimization.

Nevertheless, the multiresolution optimization has the drawback to generate “out of the range” disparities, i.e., disparity values larger than the maximum disparity (MaxDisp in Fig. 1), by instance larger than 60 pixels. Since the searching zone includes the four candidates plus the tolerance zone if the candidate point is always a point at the same edge of the searching zone (i.e., belonging to the tolerance zone) for all levels of the pyramid, then the matching point at level 0 can be separated from the reference point by a distance larger than the maximum disparity.

Fig. 15 Dense disparity map generation using the locally scaled sum of absolute differences stereo matching method. (a) Left stereo image extracted from the virtual dataset. (b) Generated disparity map.

Table 3 Mean, standard deviation, minimum, and maximum of the error scores computed for the LSAD stereo matching method over 130 points on average per image type for the three different image types (i.e., real, semivirtual, and virtual) at level 0 of the pyramid (i.e., high resolution). The mean, the standard deviation, the minimum and the maximum are calculated for two different numbers of levels of the pyramid and two parameters, i.e., a 13×13 convolution kernel and a maximum disparity of 60. The abbreviation *no pyr* designates a stereo matching computation carried out with only one level, i.e., at the original image resolution. The abbreviation *pyr* denotes a computation performed with a multiscale pyramid comprising seven levels.

Method: LSAD	Level 0			
	Mean	Std	Min	Max
Real dataset—no pyr	0.3451	0.1744	0.0109	0.675
Real dataset—pyr	0.0348	0.0402	0	0.2112
Semivirtual dataset—no pyr	0.3298	0.1922	0.0065	0.8923
Semivirtual dataset—pyr	0.0222	0.0218	0	0.1968
Virtual dataset—no pyr	0.2033	0.1364	0.0041	0.5259
Virtual dataset—pyr	0.0563	0.0524	0	0.1662

Therefore, the multiresolution optimization can produce wrong disparities, i.e., errors.

Several disparity maps were produced for the three different image types using the LSAD method, a 13×13 convolution kernel and a pyramid with seven levels. Points with disparities out of range are spread over the disparity map. Figure 16 is a zoom of a disparity map of a pair of stereo x-ray breast images of

Fig. 16 Zoom of a disparity map of a pair of stereo x-ray breast images of the virtual dataset highlighting disparities out of range.

the virtual dataset with highlighted out-of-range disparities. The number of points with disparities out of range was always less than 0.1% of the total number of pixels. Therefore, we consider the errors generated by the pyramid as being negligible. When such points are found, the returned disparity is set to a default value and thus should be considered as a hole to be filled by some postprocessing.

Finally, the multiresolution optimization is used to generate dense disparity maps from stereo x-ray breast images made of more than 16.5 million pixels. The C++ framework is optimized but other optimizations such the parallelization can be added to speed up the processing.

5.3 Local Versus Global Stereo Matching Methods

The LBP global method is compared with the eight aforementioned local methods. Due to constraints inherent to the LBP implementation, the global method is applied to the three types of images but at a resolution lower than the resolution of origin. Consequently, for the eight local methods, the mean, standard deviation, minimum, and maximum of all the error scores were computed over the three different image types at level 3 of the pyramid. Results depicted in Table 4 suggest that although the LBP method performs better than the ZNCC and the NSSD local methods, the LBP method still has a matching error much larger (i.e., 13%) than the best error values (e.g., 4.75% of matching errors for the LSAD method). Therefore, the LBP global method can be excluded.

Table 4 Mean, standard deviation, minimum, and maximum of all the error scores. For the eight local stereo matching methods, the aforementioned data are computed over the three different image types at level 3 of the pyramid, still considering a 13×13 convolution kernel and a maximum disparity of 60. For the LBP global method, mean, standard deviation, minimum, and maximum of all the error scores are computed over the three different image types but for an image resolution lower than the resolution of origin. Thus, the maximum disparity is adjusted accordingly, i.e., to 7 ($60/2^3$).

Level 3					
n#	Methods	Mean	Std	Min	Max
1	GT	0	0	0	0
2	SAD	0.0600	0.0561	0	0.2853
3	SSD	0.0572	0.0541	0	0.2694
4	ZSAD	0.0496	0.0470	0	0.2267
5	ZSSD	0.0494	0.0413	0	0.2126
6	LSAD	0.0475	0.042	0	0.1905
7	LSSD	0.0493	0.0411	0	0.2126
8	ZNCC	0.1460	0.1167	0	0.4747
9	NSSD	0.3604	0.1063	0.0999	0.5709
10	LBP_7	0.1301	0.1120	0	0.4707

Note: Bold value denotes the result of the method with the best performance (per column).

6 Discussion

Although LSAD based stereo matching method appears to be a good candidate for stereoscopic breast x-ray images with an error score of 0.038 on average, other methods such as ZSAD, ZSSD, and LSSD perform approximately the same with an error score of 0.04 on average for the three different image types (i.e., real, semivirtual, and virtual image types). In the future, newly created stereo matching methods either dedicated to x-ray breast images or expected to be efficient for this type of images can be included into the framework and compared to the aforementioned stereo matching techniques.

The error of 3.8% can be partly caused by an inherent error introduced during the manual placement of the matched points during the ground truth generation. It is sometimes difficult to find the exact position of the matching points in the right image, and therefore the manually generated ground truth may not be perfect. If a stereo matching technique performs extremely well, it may provide better stereo matchings than the ground truth, resulting in a non-null error score regarding the performance metric (defined in Sec. 4.3.4). For future development and comparison of new stereo matching techniques, the best stereo matching methods may be used for generating the ground truth in a semiautomatic mode. Then, a visual validation will still be required. Another approach to create a perfect ground truth consists in knowing upfront the matching points, e.g., during the creation of the virtual dataset when the full geometry is known. Unfortunately, we do not have access to the ground truths of the virtual dataset.

Occlusions are reference points for which not a single candidate has been found (i.e., points in the right image for which the costs computed with the reference left point are equals). The number of occlusions is very small (i.e., less than 0.1% of the total number of pixels). This number does not represent the amount of errors, i.e., wrong stereo matching, made by the framework. The amount of errors, i.e., wrong stereo matching, was estimated in Table 1 as being about 3.8% for the LSAD method. The occlusion number means that for one reference point, the framework (almost) always finds one single matching point (i.e., one candidate). This is a limitation of our framework and a potential improvement could be to consider as occlusions the points with poor stereo matching, i.e., reference points for which the cost with their candidates is higher than a dedicated “good matching” threshold.

Finally, as mentioned in Sec. 4.3, the framework can be used not only to compare results of stereo matching methods with a ground truth but also to generate dense disparity maps from the stereo x-ray breast images.

7 Conclusion

A framework for comparing stereo matching methods with x-ray stereoscopic breast images has been developed. Eight local methods have been evaluated and four of them, i.e., LSAD, ZSAD, ZSSD, and LSSD, appear to perform equally good with an average error score of 0.04. The framework can be used to generate dense disparity maps as well. The multiresolution optimization is used to generate dense disparity maps from stereo x-ray breast images made of more than 16.5 million pixels. Finally, additional stereo matching methods expected to be good candidates regarding x-ray breast image can be included in the framework and compared to the previously tested methods.

Acknowledgments

This work has been supported by the IWT (Institute for the Promotion of Innovation by Science and Technology in Flanders) in the context of the Baekeland Grant “Optical simulation, modeling and evaluation of 3-D medical displays,” 110542. Special thanks goes to Dr. Predrag Bakic from the University of Pennsylvania for his support and for providing the virtual datasets.

References

1. D. Getty, C. D’Orsi, and R. Pickett, “Stereoscopic digital mammography: improved accuracy of lesion detection in breast cancer screening,” *Digital Mammography*, E. Krupinsky, Ed., pp. 74–79, Springer Berlin Heidelberg, Heidelberg, Germany (2008).
2. C. D’Orsi et al., “Stereoscopic digital mammography: improved specificity and reduced rate of recall in a prospective clinical trial,” *Radiology* **266**, 81–88 (2013).
3. N. J. Bonnie, “Advance in breast imaging: mammography and much more,” *J. San Francisco Med. Soc.* **88**(2), 20–21 (2015).
4. B. Tutt, “New breast imaging modalities show promise for cancer screening and staging,” *OncoLog* **60**(6), 1–3 (2015).
5. A. Lohre et al., “Initial result of a prospective study: comparison between a low dose 3D stereo mammography and FFDM,” in *Breast Imaging*, A. D. A. Maidment, P. R. Bakic, and S. Gavenonis, Eds., Vol. **7361**, pp. 583–588, Springer Berlin Heidelberg, Heidelberg, Germany (2012).
6. L. T. Niklason et al., “Digital tomosynthesis in breast imaging,” *Radiology*, **205**(2), 399–406 (1997).
7. G. S. Muralidhar, A. C. Bovik, and M. K. Markey, “Machine learning in computer-aided diagnosis: medical imaging intelligence and analysis: medical imaging intelligence and analysis,” Chapter in *Computer-*

- Aided Detection and Diagnosis for 3D X-Ray Based Breast Imaging*, pp. 66–85, IGI Global (2012).
8. A. D. A. Maidment, P. R. Bakic, and M. Albert, "Is stereomammography possible without increasing dose?" in *Digital Mammography*, H.-O. Peitgen, Ed., pp. 510–516, Springer Berlin Heidelberg, Heidelberg, Germany (2003).
 9. A. D. Maidment, P. R. Bakic, and M. Albert, "Effects of quantum noise and binocular summation on dose requirements in stereoradiography," *Med. Phys.* **30**(12), 3061–3071 (2004).
 10. R. A. Castellino, "Computer aided detection (CAD): an overview," *Cancer Imaging* **5**(1), 17–19 (2005).
 11. A. Jalalian et al., "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clin. Imaging* **37**(3), 420–426 (2013).
 12. R. Murakami et al., "Detection of breast cancer with a computer-aided detection applied to full-field digital mammography," *J. Digital Imaging* **26**(4), 768–773 (2013).
 13. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision* **47**(13), 7–42 (2002).
 14. S. Akkoul et al., "3D femur reconstruction using x-ray stereo pairs," in *Image Analysis and Processing—ICIAP 2013*, A. Petrosino, Ed., Vol. **8157**, pp. 91–100, Springer Berlin Heidelberg, Heidelberg, Germany (2013).
 15. D. M. Chelberg et al., "Digital stereomammography," in *Proc. of the 2nd Int. Workshop on Digital Mammography*, pp. 181–190 (1994).
 16. D. Scharstein, R. Szeliski, and H. Hirschmüller, "Middlebury stereo vision page," 3 June 2015, <http://vision.middlebury.edu/stereo/> (22 October 2015).
 17. T. Kanade et al., "Development of a video-rate stereo machine," in *Proc. 1995 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems 95. Human Robot Interaction and Cooperative Robots*, Vol. 3, pp. 95–100, IEEE (1995).
 18. H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(9), 1582–1599 (2009).
 19. G. Abdulfattah and M. Ahmad, "Face localization-based template matching approach using new similarity measurements," *J. Theor. Appl. Inf. Technol.* **57**(3), 424–431 (2013).
 20. S. Chambon and A. Crouzil, "Combination of correlation measures for dense stereo matching," in *Int. Joint Conf. on Computer Vision Theory and Applications* (2011).
 21. L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Comput. Vision* **3**(3), 209–238 (1989).
 22. T. Ryan, R. Gray, and B. Hunt, "Prediction of correlation errors in stereo-pair images," *Opt. Eng.* **19**(3), 193312 (1980).
 23. J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artif. Intell.* **29**(3), 241–288 (1986).
 24. S. Mukherjee and R. Guddeti, "A hybrid algorithm for disparity calculation from sparse disparity estimates based on stereo vision," in *2014 Int. Conf. on Signal Processing and Communications (SPCOM)* (2014).
 25. J. Banks et al., "A hybrid stereo matching algorithm incorporating the rank constraint," in *1999. Seventh Int. Conf. on Image Processing and its Applications*, Vol. 1, IET (1999).
 26. N. Slesareva, A. Bruhn, and J. Weickert, "Is Stereomammography Possible Without Increasing Dose?," *Pattern Recognition*, W. G. Kropatsch, R. Sablatnig, and A. Hanbury, Eds., Vol. **3663**, pp. 33–40, Springer Berlin Heidelberg, Heidelberg, Germany (2005).
 27. P. Bakic et al., "Mammogram synthesis using a three-dimensional simulation. iii. Modeling and evaluation of the breast ductal network," *Med. Phys.* **30**, 1914–1925 (2003).
 28. C. Zhang, P. Bakic, and A. Maidment, "Development of an anthropomorphic breast software phantom based on region growing algorithm," *Proc. SPIE* **6918**, 69180V (2008).
 29. K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Mag.* **2**(11), 559–572 (1901).

Johanna Rousson received her MS degree in optics and photonics from Télécom-Saint-Etienne, France, in 2011. Since 2012, she has been a PhD student affiliated to both Barco NV, Belgium, and the Ghent University, Belgium. Her main research interests involve optical simulation, modeling, and evaluation of 3-D medical displays. She is a student member of SPIE and a member of SID.

Mathieu Naudin is a PhD student at the XLIM-SIC laboratory of the University of Poitiers, France, in collaboration with SIEMENS and the University Medical Center of Poitiers. He received his MS degree in biology and computer sciences in 2014 and his BS degree in biology in 2012 from the University of Poitiers, France. His research interests include medical imaging and medical decision aids. He contributed to the reported study during an internship within Barco NV.

Cédric Marchessoux received his PhD in electronics engineering from the University of Poitiers, France, in 2003. After a postdoctorate at Agfa-Gevaert, Belgium, he joined in 2006 the Barco Healthcare Division, Kortrijk, Belgium. He is the coauthor of more than 60 publications in journals and conferences. He is the coauthor of several patents and reviews for different scientific journals. Since 2004, he has been involved in several international and European projects.