# Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment

Muhammad Kashif *, Thomas M. Deserno, Daniel Haak, Stephan Jonas

*Department of Medical Informatics, RWTH Aachen University, Pauwelsstr. 30, 52057 Aachen, Germany*

A B S T R A C T

Solving problems in medical image processing is either generic (being applicable to many problems) or specific (optimized for a certain task). For example, bone age assessment (BAA) on hand radiographs is a frequent but cumbersome task for radiologists. For this problem, many specific solutions have been proposed. However, general-purpose feature descriptors are used in many computer vision applications. Hence, the aim of this study is (i) to compare the five leading keypoint descriptors on BAA, and, in doing so, (ii) presenting a generic approach for a specific task. Two methods for keypoint selection were applied: sparse and dense feature points. For each type, SIFT, SURF, BRIEF, BRISK, and FREAK feature descriptors were extracted within the epiphyseal regions of interest (eROI). Classification was performed using a support vector machine. Reference data (1101 radiographs) of the University of Southern California was used for 5-fold cross-validation. The data was grouped into 30 classes representing the bone age range of 0–18 years. With a mean error of 0.605 years, dense SIFT gave best results and outperforms all published methods. The accuracy was 98.36% within the range of 2 years. Dense SIFT represents a generic method for a specific question.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Problems in computer science and especially computer vision can be tackled by two different approaches: (i) a specific solution that utilizes a lot of prior knowledge or (ii) a general solution that is applicable to other problems, too. Bone age assessment (BAA) is the process of determining the skeletal maturity of a person. For BAA, many specific solutions have been proposed. However, it might be efficient to solve such a problem using general algorithms.

In clinical practice, the bone age of a child (developmental age of the bones) is assessed based on a radiological examination of the left hand and wrist and compared to the chronological age. This allows anticipating the adult height as well as diagnosis and management of endocrine disorders and pediatric syndromes [1]. Moreover, BAA is used in forensic medicine [2]. Another relevant application is found in social fields. According to United Nations Children's Fund (UNICEF), only half of the children under 5 years in the developing world have their births registered. In sub-Saharan Africa and South Asia, about 65% of all births go unregistered [3]. Without documented proof of age, children are recruited to fighting forces, exposed to hazardous forms of work, forced to early marriages, and treated as adult in legal proceedings. In all of these cases, skeletal maturity can help to estimate the chronological age of a person.

However, BAA is a time-consuming and cumbersome task in radiology. In clinical routine, two methods are applied: Greulich and Pyle (GP) [4] and Tanner and Whitehouse (TW) [5]. Following the GP method, the radiologist compares all bones of the left hand to those in a radiograph of a standard atlas and assesses the bone age according to his visual perception. Following the TW method, certain subsets of bones are examined with respect to epiphyseal distances. Hence, the GP method is more subjective, while the TW method is more complex and time consuming. Based on the physicians expertize, the examination time varies. In [6], an average time of 40 s and 80 s was reported for GP and TW methods, respectively. Conversely in [7], the average reading time is 84 s and 474 s for GP and TW methods, respectively. Hence, automated BAA is desired.

Many approaches have already been adopted to automate BAA. In 1996, Al-Taani et al. presented an automatic BAA approach that is based on a point distribution model (PDM) of 130 feature points [8]. The distal and middle phalanxes of third finger were classified. A set of 120 images was used for classification. The evaluation rates for two experiments were 73.7% and 70.5%. In 2001, Pietka et al. comprehensively reviewed early approaches for BAA and presented a method for feature extraction from left hand radiograph by measuring the gap between metaphyses and diaphyses

[9]. A solid view on fundamental principles in BAA was provided but computations were not performed.

Bocchi et al. proposed a system to implement the TW method using neural networks [10]. A set of 120 images for training and a set of 40 images for testing were used. A maximum error of 1.4 years with standard deviation 0.7 was reported. BAA based on phalangeal features was presented by Chang et al. [11]. In this method, the back propagation of neural networks was applied to train the features of phalanges. A quite large error of 1.5 years was reported.

In 2007, Kim and Kim used epiphyseal regions of interest (eROIs) [12]: the discrete cosine transform and a linear discriminant analysis were applied on nine relevant eROIs. A mean error of 0.6 years was reported. The data set in use was quite small and the error rate could not be confirmed by any others.

The use of private data restricts the comparability of BAA approaches. To improve this situation, a digital hand atlas of carefully selected radiographs was released at the University of Southern California (USC) and has been established as standard reference database. First experiment on that dataset was performed by Gertych et al., where a fuzzy classifier was applied on carpal bone and phalangeal ROIs [13].

In our previous work, a method based on content-based image retrieval (CBIR) was presented, where eROIs patches were extracted automatically and similar patches were retrieved using the Image Retrieval and Medical Application (IRMA) framework [14]. Classification was done with a k-nearest neighbor (kNN) approach. A mean error of 0.97 years for the age range of 0–18 years was reported on the USC data. The method was extended by Harmsen et al. introducing class prototypes. Applying the support vector machine (SVM) for classification, a mean error of 0.83 years was achieved [15]. Haak et al. have improved to 0.768 years by replacing the SVM with a support vector regression (SVR) [16]. To obtain the features, cross correlation between test and reference images was used [14–16].

The leading commercial product for BAA (BoneXpert) applies an active shape model [17]. Within the bone age ranges of 2.5–17 years and 2–15 years, BoneXpert obtains a root mean square error of 0.61 years for boys and girls, respectively [18].

However, all BAA methods published so far are specific rather than generic. In the last decade, many robust methods to extract distinct features from the image have been presented, which are being used in large variety of computer vision applications. The most prevailing methods are scale invariant feature transform (SIFT), speeded up robust feature (SURF), binary robust independent elementary features (BRIEF), binary robust invariant scalable keypoints (BRISK), and fast retina keypoint (FREAK). A lot of research has been published to compare these methods (Table 1), but a superior method has not yet been identified in general. Rather, the performance of the methods depends on the application domain.

The process of feature extraction is composed of feature detection and feature description. In feature detection, an algorithm determines the appropriate keypoints that represent the most informative part of the image. In feature description, a local image descriptor is computed for every keypoint. The descriptor possesses the neighborhood information of the keypoint to identify the same keypoint across various images.

The majority of previous research is concentrated on the comparison of feature detectors rather than feature descriptors. For instance, Juan and Gwun compared the feature detection performance of SIFT, PCA-SIFT and SURF methods for scale, rotation, and affine transforms as well as for blur and illumination changes [19]. SIFT performed superior in all experiments but showed the longest processing times. In other experiments, SURF was found fastest and stable. PCA-SIFT performed good in rotation and illumination changes.

Tuytelaars and Mikolajczyk presented a survey on local invariant feature detectors [20]. They compared corner, blob, and region detectors. Again, the study was focused on feature detectors.

Canclini et al. evaluated the performance of feature detectors and descriptors in terms of processing time, repeatability, and matching accuracy for image retrieval application [21].

Several feature extractors were compared for visual simultaneous localization and mapping (VSLAM). Klippenstein and Zhang have compared the Harris detector, the Lucas–Kanade–Tomasi tracker, and the detector part of SIFT for VSLAM [22]. They concluded that the choice of feature detector is irrelevant in terms of VSLAM performance. However, feature descriptors were not evaluated. More recently, Hartmann et al. have evaluated the feature descriptors for accuracy and speed in a typical graph-based VSLAM algorithm [23].

Nevertheless, the appropriate choice of SIFT, SURF, BRIEF, BRISK, or FREAK cannot be answered yet, since the proof of the pudding is in the eating. Moreover, a comparison of feature descriptors with respect to a certain application is not yet presented. From an evaluation-methodology point of view, a well-defined reference problem and a large, public available database of ground truth is required. Furthermore, the application domain shall be researched comprehensively on that database. Therefore, we selected the BAA problem to analyze SIFT, SURF, BRIEF, BRISK and FREAK methods for feature description rather than extraction.

## 2. Material and methods

The image processing chain in this work is composed of several steps: eROIs extraction, feature points specification, features description, and classification (Fig. 1).

### 2.1. EROI Extraction

From prior work, a semi-automatic approach is used to extract the eRIOs from radiographs [14]. For proper localization, the user simply clicks into the centers of relevant epiphyses. Fourteen eROIs were extracted from each radiograph and rotated into a normalized upright position (Fig. 2).

### 2.2. Feature point specification

Except BRIEF, all other methods (SIFT, SURF, BRISK, FREAK) can be used for keypoints detection, too. SIFT, SURF, BRISK and FREAK

**Table 1**
Feature extraction methods compared by previous authors.

|      | SIFT | PCA-SIFT | SURF | BRIEF | ORB | BRISK | FREAK | Others |
|------|------|----------|------|-------|-----|-------|-------|--------|
| [19] | x    | x        | x    |       |     |       |       |        |
| [20] | x    |          | x    |       |     |       |       | x      |
| [21] | x    |          | x    | x     | x   | x     | x     | x      |
| [22] | x    |          |      |       |     |       |       | x      |
| [23] | x    |          | x    | x     | x   | x     | x     |        |