Evaluation of 3D Operators for the Detection of Anatomical Point Landmarks in MR and CT Images

Thomas Hartkens,[1] Karl Rohr,[2] and H. Siegfried Stiehl

FB Informatik, Arbeitsbereich Kognitive Systeme, University of Hamburg, Germany

Point-based registration of images strongly depends on the extraction of suitable landmarks. Recently, different 3D operators have been proposed in the literature for the detection of anatomical point landmarks in 3D images. In this paper, we investigate nine 3D differential operators for the detection of point landmarks in 3D MR and CT images. These operators are based on either first, second, or first and second order partial derivatives of an image. In our investigation, we use measures which reflect different aspects of the detection performance of the operators. In the first part of the investigation, we analyze the number of corresponding detections under elastic deformations and noise, in the second part, we use statistical measures to determine the detection performance for landmarks within regions of interest (ROIs), and in the third part, we investigate the separability of the detections. It turns out that operators based on only first order partial derivatives of an image (i) yield a larger number of corresponding points than the other operators, (ii) that their performance on the basis of the statistical measures is better, and (iii) that the separability of the detections is better so that a suitably chosen threshold can significantly decrease the number of false detections. © 2002 Elsevier Science (USA)

1. INTRODUCTION

The registration of tomographic images is important for diagnosis and surgery planning. One possibility to match two images is to find corresponding points and to use them for registration. As corresponding points we here consider point landmarks in tomographic images, i.e., prominent points, where the surface of anatomical structures is strongly curved, e.g., the tip of the frontal horn of the ventricular system within the human brain. Usually,

[1] Current affiliation: Division of Radiological Sciences, Guy's Hospital, King's College London, London, UK.

[2] Current affiliation: School of Information Technology, International University in Germany, Bruchsal, Germany.

such 3D landmarks are selected manually—a task which is tedious, time-consuming, and often lacks accuracy. An alternative is a semi-automatic procedure for landmark extraction which has the advantage that the user can interactively control the results. First, a coarse position of a certain landmark is determined manually. Second, a 3D operator is applied within a region of interest (ROI) centered at the coarse position to detect potential landmark candidates. Third, the user selects the most promising candidate. While the *localization* performance of 3D operators has already been investigated (e.g., Frantz *et al.* [9]), studies on the *detection* performance of 3D operators can hardly be found.

In this paper, we experimentally investigate nine 3D differential operators for the detection of 3D point landmarks in MR and CT images. These operators are based on either first, second, or first and second order partial derivatives of an image. The main questions of our investigation are (1) which operators yield the largest number of corresponding points, (2) which operators detect the landmarks most reliably, and (3) which operators yield the best distribution of the operator responses, such that the application of a suitable threshold can significantly decrease the number of false detections while keeping the number of correct detections. To answer these questions, we introduce quantitative measures which represent different aspects of the detection performance. First, we determine the number of corresponding points in images under elastic deformations and noise. Second, we use statistical measures to determine the detection performance for landmarks within ROIs. Third, we analyze the distribution of the operator responses and apply a measure of separability to investigate whether the largest operator response can well be separated from the other responses. We present experimental results for 3D synthetic and 3D tomographic images. The nine 3D operators are applied for 22 3D synthetic images and 43 ROIs of 3D tomographic images which are extracted from four 3D MR images and one 3D CT image. Additionally, these images are disturbed with 10 different levels of additive Gaussian noise and are elastically deformed three times on the basis of thin-plate splines. In total, we use 308 ROIs in synthetic images and 301 ROIs in tomographic images.

Alternative studies on the performance of 3D landmark operators are based on the number of matched points under rigid transformations [21] or determine the rigid or affine registration accuracy [2, 21]. Also, these studies are less comprehensive and only a relatively small number of operators has been considered. For evaluation studies of 2D operators see, e.g., [5, 10, 12, 19, 22]. The current evaluation study for the detection performance of 3D operators is more comprehensive than our previous one [11] and additionally includes an investigation of the separability of the detections.

The organization of this paper is as follows: First, we briefly describe the nine investigated differential operators (Section 2). Then, we introduce measures for three aspects of the detection performance (number of corresponding points, statistical measures, and a measure characterizing the separability of the operator responses), which we use to compare the operators (Section 3). The parameter settings applied in our study as well as the used image data are described in Section 4. The experimental results are presented in Sections 5–8, which also include an analysis of the operator responses for 3D synthetic and 3D tomographic images.

2. INVESTIGATED 3D OPERATORS

We have investigated nine 3D differential operators for detecting anatomical point landmarks in 3D images $g(x, y, z)$ as summarized in Table 1. All operators are differential operators and are based on partial derivatives of an image up to second order. Since most

TABLE 1

The Investigated Nine 3D Differential Operators Classified into Four Different Approaches

Mean curvature approach [3, 7, 13, 16, 20]

- $H = \dfrac{1}{2|\nabla g|^3}\Big[g_x^2(g_{yy}+g_{zz}) + g_y^2(g_{xx}+g_{zz}) + g_z^2(g_{xx}+g_{yy}) - 2(g_x g_y g_{xy} + g_x g_z g_{xz} + g_y g_z g_{yz})\Big]$
- $Kitchen\ \&\ Rosenfeld3D = H \cdot 2|\nabla g|$
- $Blom3D = H \cdot 2|\nabla g|^3$ with $\nabla g = (g_x, g_y, g_z)^T$

Gaussian curvature approach [2, 7, 20]

- $K = \dfrac{1}{|\nabla g|^4}\Big[g_x^2 \cdot \big(g_{yy}g_{zz} - g_{yz}^2\big) + 2g_y g_z \cdot (g_{xz}g_{xy} - g_{xx}g_{yz}) + g_y^2 \cdot \big(g_{xx}g_{zz} - g_{xz}^2\big)$
  $\quad + 2g_x g_z \cdot (g_{yz}g_{xy} - g_{yy}g_{xz}) + g_z^2 \cdot \big(g_{xx}g_{yy} - g_{xy}^2\big) + 2g_x g_y \cdot (g_{xz}g_{yz} - g_{zz}g_{xy})\Big]$
- $K^* = K \cdot |\nabla g|^4$

Förstner–Rohr approach [8, 14, 16]

- $Op3 = \dfrac{det(\underline{C})}{trace(\underline{C})}.$
- $Rohr3D = det(\underline{C})$
- $Foerstner3D = \dfrac{1}{trace(\underline{C}^{-1})} = \dfrac{det(\underline{C})}{trace(\underline{C}^{adj.})}$ with $\underline{C} = \overline{\nabla g(\nabla g)^T}$

Beaudet approach [1]

- $Beaudet3D = det(H_g)$ H_g Hessian matrix

of these operators are 3D extensions of 2D corner operators we denote them by their corresponding authors who introduced the 2D operators, together with the suffix 3D. Three of the nine operators are based on the mean curvature H of isocontours (H, *Kitchen&Rosenfeld3D*, and *Blom3D*) [3, 7, 13, 16, 20] and two operators are based on the Gaussian curvature K (K and K^*) [2, 7, 20]. According to one approach the operators (based on either H or K) differ only by the exponent of the gradient magnitude. For example, the operator *Blom3D* results from multiplying the operator *Kitchen&Rosenfeld3D* by the second power of the gradient magnitude. Another three operators (*Op3*, *Rohr3D*, and *Förstner3D*) [8, 14, 16] are based on the matrix $\underline{C} = \overline{\nabla g(\nabla g)^T}$ with $\nabla g = (g_x, g_y, g_z)^T$ and have been summarized under the Förstner–Rohr approach. Note that the latter three operators require only first order partial derivatives of an image. One operator exploits the Hessian matrix H_g and includes only second order partial derivatives (operator *Beaudet3D*) [1].

3. MEASURES FOR THE DETECTION PERFORMANCE

To compare the detection performance of the operators we use different measures. We investigate the number of corresponding points and statistical measures as well as the separability of the operator responses.

3.1. Number of Corresponding Points

The detected points should be both invariant w.r.t deformations and robust against noise. Thus, we should detect the same points if we deform the images or add noise to the images. In our study we investigate the stability w.r.t. deformations by deforming the images with an elastic transformation [4, 18] and determining automatically the corresponding points

in the original image and the deformed image. The elastic transformation is determined by a set of point pairs, while corresponding positions are obtained by applying a random generator. To reduce the dependence on a specific deformation, we deform the images three times and count the number of corresponding points in all of these images. To study the stability w.r.t. noise we investigate three different levels of Gaussian noise.

3.2. Statistical Measures

We use three different statistical measures to investigate how reliably the operators detect landmarks. For each landmark within a ROI (e.g., $25 \times 25 \times 25$ voxels) we use a detection region ($7 \times 7 \times 7$ voxels) which has the advantage that small localization errors of the operators (cf. [9, 15]) do not distort the detection performance. If at least one detection is within this detection region, we consider the landmark to be detected and the detected point to be a *correct detection*. If no detection is inside the detection region, we have a *misdetection* (false negative), whereas if more than one detection is inside the detection region, the landmark is *multiple detected*. A detected point outside the detection region is called a *false detection* (false positive). After application of the operators we determine the overall number of detections (n_d), the number of correct detections ($n_{d,in}$), and the number of detected landmarks ($n_{l,detect}$), which represents the number of landmarks that are detected at least once. The overall number of landmarks is denoted by n_l. Based on these quantities we compute the following measures of the detection performance,

$$P_{in} = \frac{n_{d,in}}{n_d}, \qquad P_{detect} = \frac{n_{l,detect}}{n_l}, \qquad P_{multiple} = \frac{n_{d,in}}{n_l}, \qquad (1)$$

which quantify the fraction of correct detections, the fraction of detected landmarks, and the average number of multiple detections per landmark, respectively. Previously, statistical measures have been applied in the case of 2D corner operators (Zuniga and Haralick [22]). There, only two measures were employed and detection regions around corners were not considered. Thus, the resulting detection performance in that work depends more strongly on the localization accuracy. Using the measures from above we can compute other measures as well, e.g., the average number of detections per landmark:

$$\frac{P_{multiple}}{P_{in}} = \frac{\frac{n_{d,in}}{n_l}}{\frac{n_{d,in}}{n_d}} = \frac{n_d}{n_l}.$$

3.3. Separability of the Operator Responses

Although the inclusion of a threshold cannot improve the number of correct detections, an appropiate threshold may decrease the number of false detections and eventually the number of multiple detections. Since the purpose of a threshold is to separate the correct detections from the false detections, a prerequisite is the separability of the operator responses, i.e., if the operator values are distributed in such a way that the choice of a certain threshold value is not critical to separate correct and false detections. Thus, the operator should yield a significantly larger response at the landmark in comparison to the other detections. We hereby assume that the largest operator response results from the considered landmark. In

this study we use the following measure of the separability of the operator responses,

$$\psi = \sum_{i=1}^{n} \frac{|R_i|}{|R_{max}|},$$

where n is the number of detections, $|R_i|$ is the absolute value of the operator response of the ith detection, and $|R_{max}|$ is the absolute value of the largest operator response (i.e., $\forall i = 1 \ldots n : |R_i| \leq |R_{max}|$) [17]. If $n = 0$, we set $\psi = 0$. If there is only one detection we obtain $\psi = 1$, and additional false detections with low operator responses yield a value of $\psi \approx 1$. In this case, the correct detection can clearly be distinguished from the false detections. On the other hand, if there are operator responses with values similar to the maximal operator response, then ψ is much larger than 1 with a maximum of n (in the case if n detections with identical operator responses occur). Thus, in summary we have here a scalar quantity which characterizes the separability of the operator responses. Note that a larger value of ψ corresponds to an increase of inseparability.

4. PARAMETER SETTINGS AND IMAGE DATA

The partial derivatives of the images are estimated by applying 3D extensions of the 2D filters of Beaudet [1] with size $5 \times 5 \times 5$ voxels. The components of the matrix $\underline{C}$ of the Förstner–Rohr approach [8, 14, 16] are determined by averaging the first order partial derivatives within an observation window of $3 \times 3 \times 3$ voxels. Extrema of the operator responses are determined by a local maximum and minimum search in $5 \times 5 \times 5$ neighborhoods.

To reduce operator responses which obviously result from noise we use a low threshold. The threshold value is chosen for each operator and each ROI separately, and those operator responses which are smaller than 1% of the maximum response of this operator given this ROI are ignored. For example, if an operator yields a maximum response of 300 in a ROI, only responses less then 3 are ignored. Since such extremely small operator responses cannot be considered as significant, it can well be assumed that the results do not depend on the chosen 1% threshold. For the whole investigation we used the same threshold value. Our investigation focuses on studying the influence of noise and deformation on the dectection performance rather than tuning the threshold parameter (for investigations on the parameter sensitivity of the threshold in the case of evaluating 2D edge detectors, see [6]).

The operators are applied to 3D synthetic and 3D tomographic images. The 3D synthetic images model either tetrahedra (aperture angles $\beta = 30°$, $40°$, $50°$, $60°$, $70°$, $80°$, $90°$), ellipsoids (lengths of the half axes $a = 8$, 9, 10, 11, 12, 13, 14, 15, 16, $b = 8$, $c = 40$), or hyperbolic paraboloids (parameters $(a, b) = (1, 1), (2, 2), (3, 3), (1, 2), (1, 3), (2, 3)$), which are smoothed by a Gaussian filter with standard deviation $\sigma = 0.7$. The tips of the tetrahedra, the tips of the ellipsoids, and the saddle points of the hyperbolic paraboloids are defined as the landmark points. In total we analyze 22 synthetic 3D images. As 3D tomographic images we use four T1 weighted MR and one CT image of the human head. The resolution of the MR images is between 0.85 and 1.0 mm, while the slice thickness is between 1.0 and 1.2 mm. The CT image has a voxel size of $0.65 \times 0.65 \times 1.0$ mm. In each of these images we consider 10 landmarks and define ROIs of size $25 \times 25 \times 25$ voxels around each landmark. As anatomical landmarks we use the tips of the frontal, occipital, and temporal horns of the ventricular system, the tip of the external occipital protuberance, the saddle point at the zygomatic bone, and the junction at the upper end of the pons. We have manually specified

the positions of these landmarks in the investigated data sets and have taken them as ground truth positions, although we know that manual localization of 3D landmarks generally is difficult and may be prone to error. Since seven of the 50 landmarks are located too close to the border of the image, in total only 43 ROIs (subimages) of the tomographic images could be used in our experiments.

5. ANALYSIS OF THE OPERATOR RESPONSES

Figure 1 shows the orthogonal sectional views of the operator responses for all nine 3D operators applied to a landmark in a synthetic image (ellipsoid) with added Gaussian noise ($\sigma_n^2 = 100$). It can be seen that the operators H and K are so sensitive to noise that the landmark (the tip of the ellipsoid) is not recognizable in the operator responses. In contrast, the operators *Kitchen&Rosenfeld3D*, *Blom3D*, and K^* yield a significantly better result. Note that these operators differ from H and K, resp., by multiplication with a certain power of the gradient magnitude. It can also been seen that for the operator *Kitchen&Rosenfeld3D* the noise influence is relatively strong everywhere in the image, while for the operators *Blom3D* and K^* the noise influence is smaller. The operator *Beaudet3D* yields two extrema at the landmark: a minimum and a maximum (see also, e.g., [15] for a similar behavior in the 2D case). The operators *Kitchen&Rosenfeld3D*, *Blom3D*, K^*, and *Beaudet3D* detect not only the landmark, but also a larger part of the surface of the ellipsoid. In comparison, for the operators of the Förstner–Rohr approach (operators *Op3*, *Rohr3D*, and *Förstner3D*), the operator responses are better concentrated at the tip and also the noise influence is smaller.

In Fig. 2 the operator responses at a landmark in a 3D MR image (tip of the left frontal horn of the ventricular system) are shown. The results are comparable to the results for the synthetic image above. In comparison to the operator *Kitchen&Rosenfeld3D* the operators *Blom3D* and K^* emphasize the surface of the ventricular system and better suppress other structures in the image. As above, the operator *Beaudet3D* yields two extrema at the landmark. The operators of the Förstner–Rohr approach yield significant values at the landmark and well suppress other structures. In particular the responses of the operators *Op3* and *Rohr3D* are better concentrated at the landmark in comparison to the other operators.
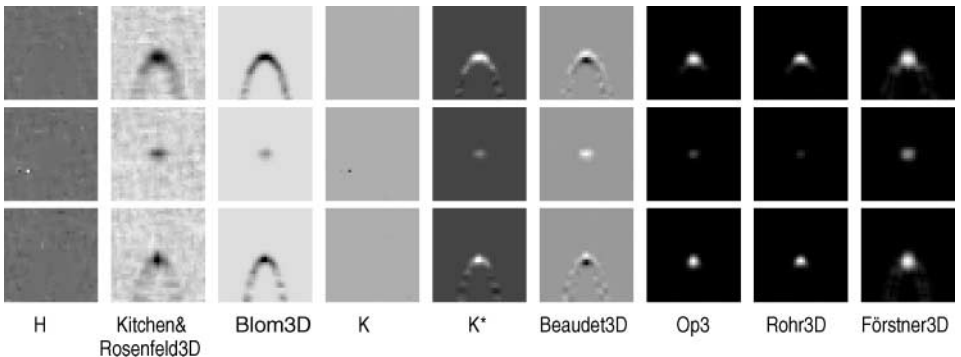

| H | Kitchen&Rosenfeld3D | Blom3D | K | K* | Beaudet3D | Op3 | Rohr3D | Förstner3D |

FIG. 1. Operator responses for an ellipsoid ($a = 10, b = 8, c = 40$) for all nine 3D operators. The figures depict in each column the orthogonal sectional views of the operator responses at the landmarks (tip of the ellipsoid). The image has been distorted with Gaussian noise of variance $\sigma_n^2 = 100$.
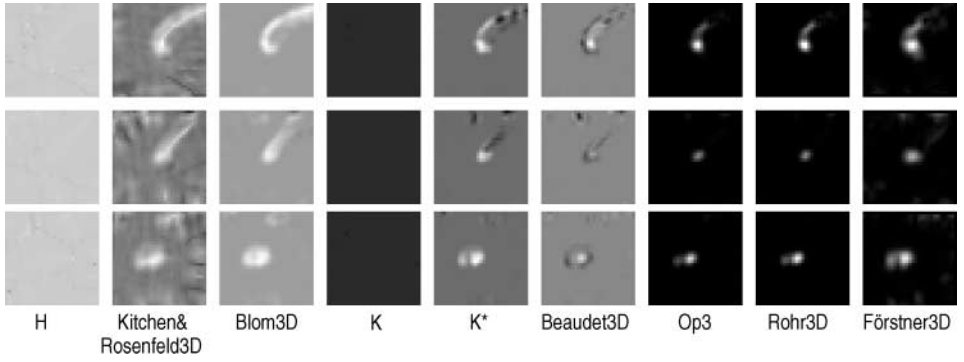
FIG. 2. Operator responses for a 3D MR image for all nine 3D operators. The figures depict in each column the orthogonal sectional views of the operator responses at the landmark frontal ventricular horn. The rows show the sagittal, axial, and coronal views for all nine 3D operators.

6. NUMBER OF CORRESPONDING POINTS

To investigate the stability of the operators w.r.t. deformations and noise, we determine the number of corresponding points in two images. First we describe the general strategy and then present the results of the investigation.

6.1. Experimental Strategy

We illustrate the general strategy for the case of deformations (see Fig. 3a). The 3D images are elastically deformed three times by applying a nonrigid transformation on the basis of thin-plate splines [4, 18]. The transformation is driven by the displacements of randomly selected points in the image, while a Gaussian noise generator is used to determine these
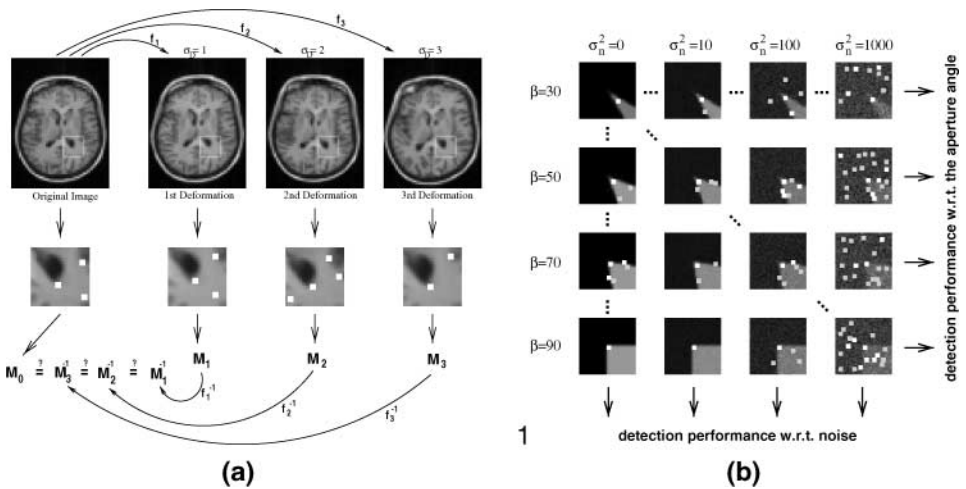

FIG. 3. General strategy for determining the number of corresponding points under elastic deformations (a) and the statistical performance measures for synthetic images (b). Fig. 3b shows the detections of the operator *Förstner3D* in 3D images of tetrahedra for different noise levels and different aperture angles. Note that the detections marked in white are within the displayed slice, while the detections marked in gray are in adjacent slices.

displacements. The chosen standard deviation of the noise is maximal 3 voxels; i.e., most points move randomly up to 3 voxels. In Fig. 3a the three transformations are denoted by f_1, f_2, and f_3. For each landmark in the original as well as in the deformed images we extract a ROI and apply the operators within these ROIs. Since the displacements of the landmarks are much smaller than the size of the ROI, we do not need to map the ROI according to the transformation, but take the same ROI in the warped image. Each operator yields sets of detected points denoted by M_0, M_1, M_2, and M_3. We transform the positions of the detections in the deformed images backwards using the inverse transformations f_1^{-1}, f_2^{-1}, and f_3^{-1} and check the fidelity of the positional correspondence of the detections in the original image. For a correct correspondence a deviation within a $3 \times 3 \times 3$ neighborhood is allowed. In the investigation of stability w.r.t. noise we add three levels of Gaussian noise to the images ($\sigma_n^2 = 1, 4, 10$) and also determine the number of corresponding points in the original and the noisy images. In this case an inverse transformation is not necessary.

A problem is that the operators generally yield a different number of detections. To make the results better comparable, we divide the number of corresponding points by the total number of detections; thus we compute the fraction of corresponding points. For each type of image (tetrahedra, ellipsoids, hyperbolic paraboloids, MR, and CT images) we average the fraction of corresponding points. Counting the different deformations and the noise levels, in total we applied all nine operators to 154 synthetic and 301 tomographic (sub)images.

6.2. Results

The diagrams in Fig. 4 depict the results of all operators applied to the 3D synthetic images with ellipsoids, to the 3D MR images, and to the 3D CT images. The diagrams in the first row show the results for the noise study, while those in the second row give the results for the case of deformations.

6.2.1. Fraction of Corresponding Points in Noisy Images

From Fig. 4 (top row, on the left) it can clearly be seen that the operators H and K yield a very small fraction of corresponding points for synthetic images with added noise.
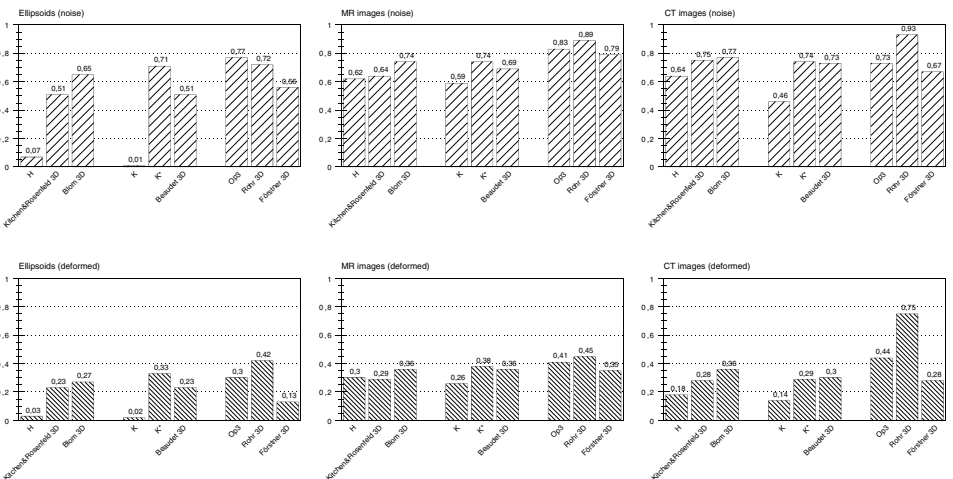

FIG. 4. Number of corresponding points (fraction w.r.t. the total number of detections).

Thus, we can conclude that these operators are very sensitive to noise, which is in accordance with the observations described in Section 5 above. The results for the MR and CT images (Fig. 4, top row, in the middle and on the right) are much better (about 60% of the detections are corresponding points). However, for the tomographic images the operators H and K yield a very large number of detections; thus, many detections seem to correspond by chance. This conjecture will be confirmed by the investigation below using the statistical performance measures (Section 7). The operators *Kitchen&Rosenfeld3D*, *Blom3D*, and K^* always detect a larger fraction of corresponding points than the operators H and K. Thus, the multiplication with the gradient magnitude (in comparision to H and K) improves the results. In particular, the operator *Blom3D* yields better results than the operator *Kitchen&Rosenfeld3D*. In all given images one of the operators of the Förstner–Rohr approach always detects the largest fraction of corresponding points. The operator *Op3* yields the best results for the images with ellipsoids whereas the operator *Rohr3D* yields the best results for the MR and CT images. In comparison to these two operators the operator *Förstner3D* performs slightly worse. The results of the operator *Beaudet3D* are worse than those of the operators *Blom3D* and K^*.

6.2.2. *Fraction of Corresponding Points in Elastically Deformed Images*

In the experiments using elastically deformed images (second row in Fig. 4) the operators yield a significantly smaller fraction of corresponding points than in the case of noisy images. However, the results for the different operators are qualitatively comparable. The operators H and K detect only a small fraction of corresponding points in the synthetic images. Multiplication with the gradient magnitude improves the results of the operators of the mean curvature (H, *Kitchen&Rosenfeld3D*, *Blom3D*) and Gaussian curvature (K, K^*) approach. In particular the fraction of corresponding points of the operator *Blom3D* is larger than that of the operator *Kitchen&Rosenfeld3D*. For the synthetic as well as the tomographic images one of the operators of the Förstner–Rohr approach always yields the best results (except for the images with a tetrahedron, for which the result for the operator K^* is better; not displayed here). In the tomographic images the operator *Rohr3D* yields the largest number of corresponding points.

7. STATISTICAL PERFORMANCE MEASURES

The analysis of the operator responses (Section 5) along with the investigation of the number of corresponding points (Section 6) showed that the influence of noise on the performance of the operators differs significantly. In this section, we use the statistical performance measures described in Section 3.2 to quantify (i) the number of correct detections P_{in}, (ii) the number of detected landmarks P_{detect}, and (iii) the average number of multiple detections $P_{multiple}$. We investigate 3D synthetic images with added Gaussian noise as well as 3D MR and CT images.

7.1. Experimental Strategy

For each type of image the statistical performance measures in (1) are investigated separately. We consider a detection region ($7 \times 7 \times 7$ voxels) around each landmark. If at least one detection is within this region, we consider the landmark to be detected and the detected

point to be a *correct detection*. The general strategy of the investigation is depicted in Fig. 3b for the case of tetrahedra. The synthetic images (e.g., for the case of tetrahedra we use aperture angles of $\beta = 30°, 40°, 50°, 60°, 70°, 80°, 90°$) have been disturbed with 10 different levels of additive Gaussian noise ($\sigma_n^2 = 0.6, 1, 2, 4, 8, 10, 50, 100, 500, 1000$). For each image we determine the measures P_{in}, P_{detect}, and $P_{multiple}$. On the one hand, we average the measures for the different parameters of the synthetic structures (e.g., in Fig. 3b we average the results over the columns) and obtain the detection performance w.r.t. these parameters. On the other hand, we average the measures over the different noise levels (e.g., in Fig. 3b we average the results over the rows) and obtain the detection performance w.r.t. the noise. For the CT and MR images we average the obtained values for the measures over all images (without added noise). In total, in this investigation all nine operators have been applied to 242 synthetic and 43 tomographic (sub)images.

7.2. Results

First, we discuss the results for the synthetic images for the case of tetrahedra (see Figs. 5 and 8; for the case of ellipsoids and hyperbolic paraboloids see Figs. 6 and 7). Then, we present the results for the MR and CT images as summarized in Fig. 9.

Synthetic images. First we consider the total number of detections (dashed line in Fig. 5) for the operators of the mean curvature approach (H, *Kitchen&Rosenfeld3D*, and *Blom3D*).
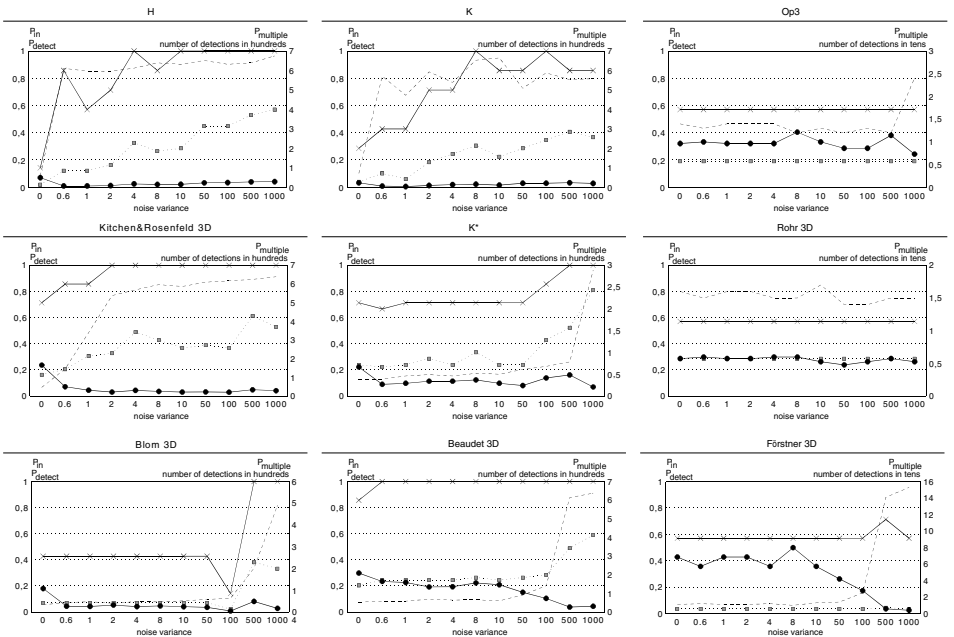

FIG. 5. Statistical performance measures for tetrahedra as a function of image noise. The measures P_{in}, P_{detect}, and $P_{multiple}$ were averaged over all different aperture angles ($\beta = 30°, 40°, 50°, 60°, 70°, 80°, 90°$) of the tetrahedra. In the diagrams these averaged measures and the number of the detections in all images are depicted. The measures P_{in} and P_{detect} refer to the units on the left side of the diagrams and the measure $P_{multiple}$ and the number of detections refer to the units on the right side of the diagrams (please note that the units have to be multiplied by 10 and 100, resp.). ●— P_{in}, ✳ P_{detect}, ▪···· $P_{multiple}$, ---- Number of detections.
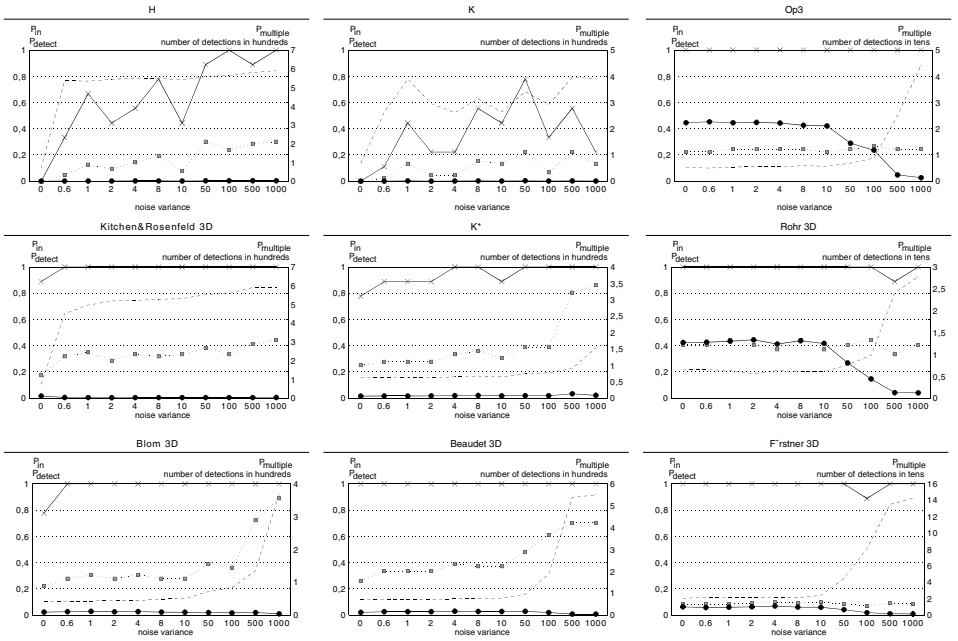
FIG. 6. Statistical performance measures for ellipsoids as a function of image noise. The measures P_{in}, P_{detect}, and $P_{multiple}$ were averaged over all lengths of the half axes ($a = 8, 9, 10, 11, 12, 13, 14, 15, 16, b = 8, c = 40$) of the ellipsoids. The units are analogous to Fig. 5. ●—P_{in}, ✳ P_{detect}, ┄■┄ $P_{multiple}$, - - - - Number of detections.
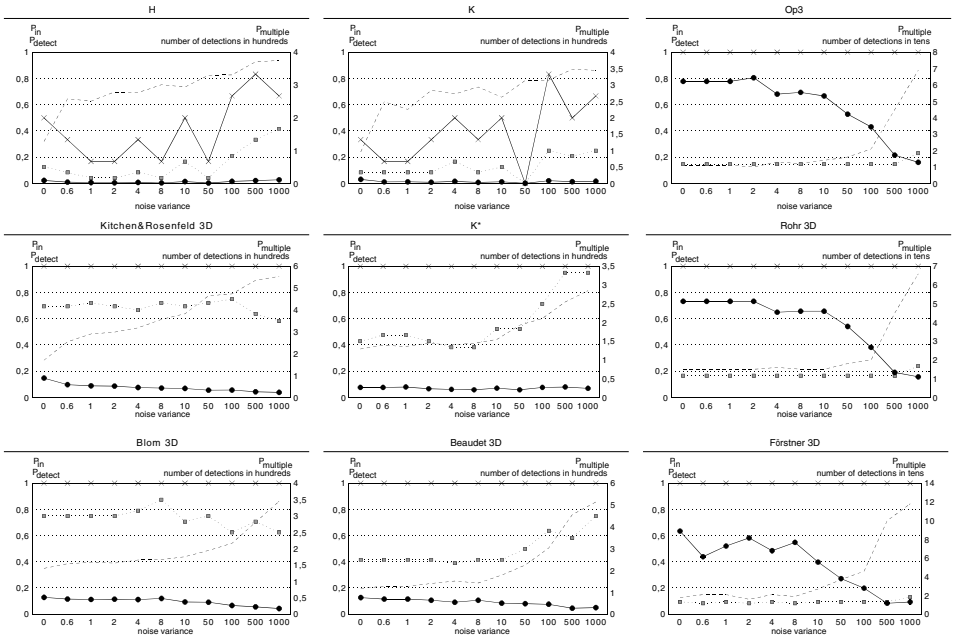

FIG. 7. Statistical performance measures for hyperbolic paraboloids as a function of image noise. The measures P_{in}, P_{detect}, and $P_{multiple}$ were averaged over all different parameters ($(a, b) = (1,1), (2,2), (3,3), (1,2), (1,3), (2,3)$) of the hyperbolic paraboloids. The units are analogous to Fig. 5. ●—P_{in}, ✳ P_{detect}, ┄■┄ $P_{multiple}$, - - - - Number of detections.
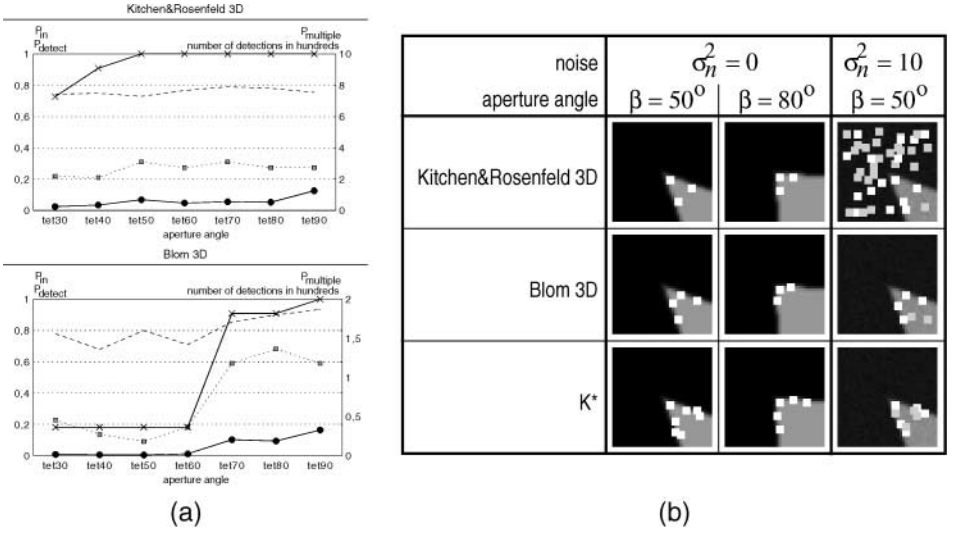
FIG. 8. Statistical performance measures for tetrahedra as a function of the aperture angle (a). In (b) the detections of three different operators in images with a tetrahedron are shown. In Fig. 8a the measures P_{in}, P_{detect}, and $P_{multiple}$ are averaged over different levels of Gaussian noise ($\sigma_n^2 = 0.6, 1, 2, 4, 8, 10, 50, 100, 500, 1000$) and are depicted as a function of the aperture angle ($\beta = 30°, 40°, 50°, 60°, 70°, 80°, 90°$). The units are analogous to Fig. 5. Fig. 8b depicts the detections of the operators *Kitchen&Rosenfeld3D*, *Blom3D*, and K^* in 3D images with and without noise and for different aperture angles $\beta = 50°$ and $\beta = 80°$. The detections marked in white are within the displayed slice, while the detections marked in gray are in adjacent slices.

For images with added noise the operator H yields a very large number of detections (about 600), which shows that this operator is rather noise sensitive. In the images without noise this operator yields a smaller number of detections, but hardly detects the landmarks ($P_{detect} < 0.2$). For the operator *Kitchen&Rosenfeld3D* the number of detections is smaller in images with low noise levels ($\sigma_n^2 \leq 1$), but for higher noise levels this operator yields about the same number of detections as the operator H. In comparison to these two operators, the operator *Blom3D* yields a much smaller number of detections in noisy images. Only in images with a high noise level ($\sigma_n^2 \geq 500$) does the number of detections significantly increase to above 200 detections. Thus, all operators of the mean curvature approach yield a large number of detections in images with high noise levels, but they differ in the noise level for which the number of detections significantly increases (H, $\sigma_n^2 \approx 0.6$; *Kitchen&Rosenfeld3D*, $\sigma_n^2 \approx 1$; *Blom3D*, $\sigma_n^2 \approx 500$). Above this noise level the operators detect all landmarks ($P_{detect} = 1$), but yield a small fraction of correct detections ($P_{in} < 0.1$). In this case, the noise influence is so large that the operators detect points everywhere in the image and thus by chance also in the detection region (e.g., see in Fig. 8b the detections of the operator *Kitchen&Rosenfeld3D* in the last column for the noise level $\sigma_n^2 = 10$). In summary, the comparison of the operators of the mean curvature approach applied to noisy images shows that the operator *Blom3D* is not as sensitive to noise as the operators H and *Kitchen&Rosenfeld3D*.

However, in the images without noise ($\sigma_n^2 = 0$) the operator *Kitchen&Rosenfeld3D* detects more landmarks than the operator *Blom3D* (see Fig. 5; *Kitchen&Rosenfeld3D*, $P_{detect} \approx 0.7$ vs *Blom3D*, $P_{detect} \approx 0.4$). To illustrate why this is the case, we consider the statistical measures as a function of the aperture angle of the tetrahedra (see Fig. 8a). While
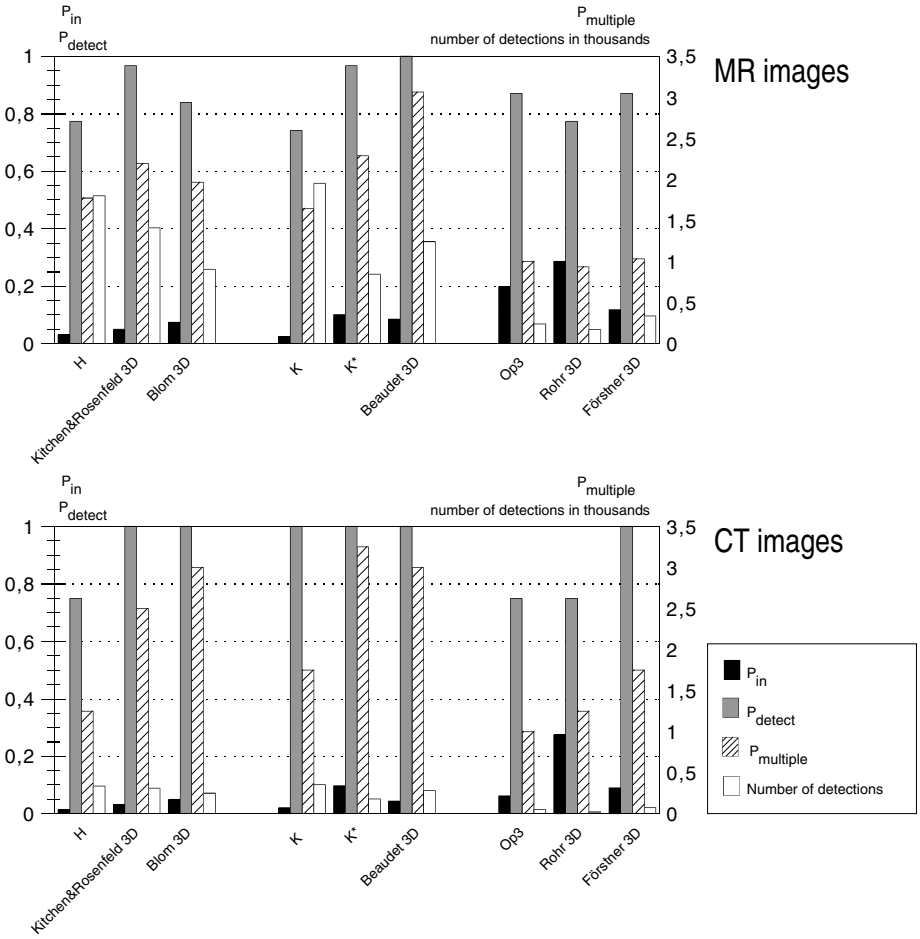
FIG. 9. Statistical performance measures for MR and CT images. The measures P_{in}, P_{detect}, and $P_{multiple}$ have been averaged over all investigated MR and CT images. Besides these measures the overall number of detections in all images has been represented in the diagrams. The measures P_{in} and P_{detect} refer to the units on the left side of the diagrams and the measure $P_{multiple}$ and the number of detections refer to the units on the right side of the diagrams.

in images with aperture angle $\beta \leq 60°$ the operator *Blom3D* hardly detects the landmarks ($P_{detect} \approx 0.2$), the operator *Kitchen&Rosenfeld3D* detects at least 70% of the landmarks ($P_{detect} \geq 0.7$). The reason is that the localization error of the operator *Blom3D* is larger than that of the operator *Kitchen&Rosenfeld3D*. For example, in the image with an aperture angle of $\beta = 50°$ and without noise the operator *Blom3D* yields no detection at the landmark (compare in Fig. 8b the detections of the operators *Kitchen&Rosenfeld3D* and *Blom3D* in the first and second columns).

We now compare the mean curvature approach with the Gaussian curvature approach by comparing the operators *H* and *Blom3D* with the operators *K* and *K**, respectively. Note that the operators *Blom3D* and *K** differ from *H* and *K*, respectively, only by multiplication with a certain power of the gradient magnitude. Figure 5 shows that the operators of the Gaussian curvature approach (*K*, *K**) yield results similar to the operators of the mean curvature approach (*H*, *Kitchen&Rosenfeld3D*, *Blom3D*). Similar to the operator *H*, the

operator K is strongly influenced by noise. Also, as with the operator *Blom3D* the noise influence of the operator K^* is smaller in comparison to the operator K (see the number of detections in Fig. 5 and the result of the operator K^* in the noisy image in Fig. 8b). However, in comparison to the operator *Blom3D* the operator K^* detects more landmarks ($P_{detect} > 0.6$) and yields a larger fraction of correct detections ($P_{in} \approx 0.1$ in noisy images). As one can see in Fig. 8b in the first row, for the image with an aperture angle of $\beta = 50°$ the localization error of the operator K^* is smaller than that of the operator *Blom3D*.

In comparison to the operators of the mean curvature and Gaussian curvature approaches, the operator *Beaudet3D* yields better results in the images with tetrahedra (Fig. 5). Both the fraction of detected landmarks P_{detect} and the fraction of correct detections P_{in} are in general significantly larger.

The operators of the Förstner–Rohr approach yield significantly more correct detections (P_{in} is larger) than the operators of all other approaches (see the last column in Fig. 5). For example, the operator *Rohr3D* yields at least 20% correct detections ($P_{in} > 0.2$) in images with and without noise. Note that the fraction of correct detections for the operators *Op3* and *Rohr3D* is hardly influenced by noise and the number of detections is at least ten times smaller than the number of detections for the operators of the other approaches. The reason why the operators of the Förstner–Rohr approach do not detect all landmarks ($P_{detect} < 1$ in Fig. 5) is that localization errors for small aperture angles of the tetrahedron occur (compare with the results of the operator *Blom3D* from above).

Tomographic images. The results for the MR and CT images are depicted in Fig. 9. Note that the measures P_{in}, P_{detect}, and $P_{multiple}$ have been averaged over all investigated MR and CT images (separately for MR and CT images). For the MR images (see Fig. 9 upper diagram) it can be seen that the operators *Kitchen&Rosenfeld3D*, K^*, and *Beaudet3D* detect more landmarks ($P_{detect} \approx 1$) than the operators of the Förstner–Rohr approach (e.g., *Op3*: $P_{detect} \approx 0.8$). However, they yield a significantly smaller fraction of correct detections (P_{in} is smaller). If both P_{in} and P_{detect} are taken into consideration, the operator *Rohr3D* yields the best results among the operators of the Förstner–Rohr approach, and the operator K^* yields the best results among the operators of the mean curvature, Gaussian curvature, and Beaudet approaches. We now analyze the results of these two operators in more detail. Since the fraction of correct detections for the operator *Rohr3D* is about 30% ($P_{in} \approx 0.3$) and since on average one detection is inside a detection region ($P_{multiple} \approx 1$), the operator detects $\frac{P_{multiple}}{P_{in}} \approx 3$ points per landmark (see Section 3.2). Thus, the operator yields per landmark one correct detection ($P_{multiple} \approx 1$) and two false detections. In comparison, the fraction of correct detections for the operator K^* is about 10% ($P_{in} \approx 0.1$) and the average number of multiple detections is $P_{multiple} \approx 2$. Thus, this operator detects about $\frac{P_{multiple}}{P_{in}} \approx 20$ points per landmark, and therefore the operator yields 18 false detections per landmark besides the two correct detections ($P_{multiple} \approx 2$). This comparison shows that the operators of the Förstner–Rohr approach (in particular the operators *Op3* and *Rohr3D*) yield a significantly smaller number of false detections. Thus it is safe to state that their detection performance is better.

A closer look at the performance measures of the operators of the mean curvature and Gaussian curvature approaches shows that the multiplication with the gradient magnitude improves the results. Both the fraction of correct detections P_{in} and the fraction of detected landmarks P_{detect} are higher. In particular, the operator *Blom3D* yields more correct detections than the operator *Kitchen&Rosenfeld3D*. Analogously, K^* performs better than K.

Note also that the operators H and K yield a relatively large number of detections in the tomographic images as we have already mentioned above in Section 6.2.1.

For the CT images the results of the different operators (Fig. 9 bottom diagram) support our findings derived from the experiments with MR images. However, the operator $Op3$ yields a smaller fraction of correct detections (P_{in} is smaller) than in the case of MR images.

8. SEPARABILITY OF THE OPERATOR RESPONSES

The investigations described above show that the operators yield in some cases a relatively large number of false detections. To separate the correct detections from the false detections a threshold can be used, provided that the largest operator response is significantly larger than the operator responses of the other detections. We investigate this condition for a noncritical threshold by applying the measure ψ as described above in Section 3.3 and use synthetic as well as MR and CT images. For the synthetic images we additionally investigate the influence of noise on the performance measure ψ.

8.1. Experimental Strategy

We disturb the synthetic images with 10 different levels of additive Gaussian noise ($\sigma_n^2 = 0$, 0.6, 1, 2, 4, 8, 10, 50, 100, 500, 1000), average the measure ψ for the different parameters of the synthetic structure, and depict the detection performance as a function of image noise (see Fig. 10). In the case of tomographic images ψ is averaged over all images (see Fig. 11). In this investigation no threshold is applied. Note that ψ measures the separability of the operator responses. In the ideal case we have $\psi = 1$ whereas for larger values of ψ the detection performance is worse.
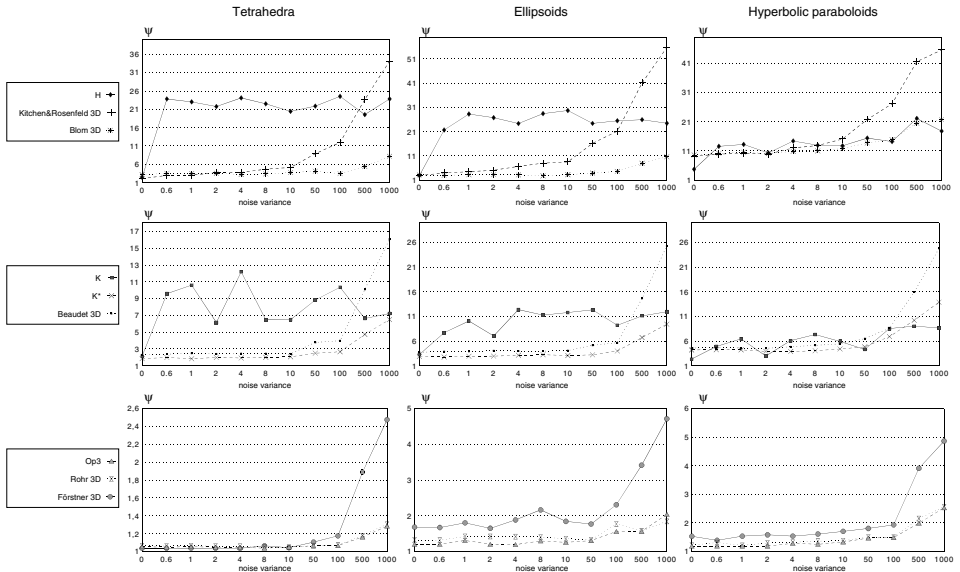

FIG. 10. Detection performance ψ for the synthetic images as a function of noise. In the columns the results for the tetrahedra, the ellipsoids, and the hyperbolic paraboloids are represented. The first row shows the results for the operators H, *Kitchen&Rosenfeld3D*, and *Blom3D*, the second row shows the results for the operators K, K^*, and *Beaudet3D*, and the third row shows the results for the operators $Op3$, *Rohr3D*, and Förstner3D.
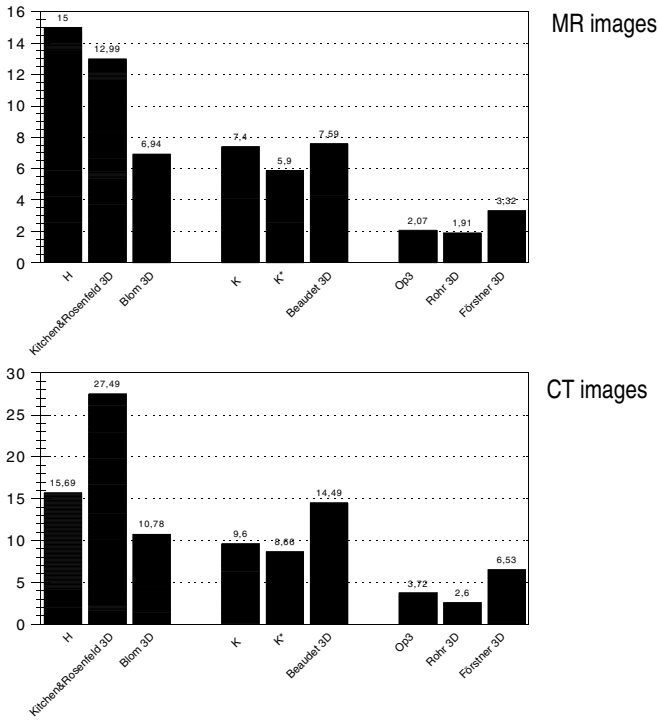
FIG. 11.   Detection performance ψ for the MR and CT images. The measure ψ has been averaged over all investigated MR and CT images.

8.2. Results

In Fig. 10 the results using the performance measure ψ for the synthetic images are depicted. First we consider the results for the operators of the mean curvature approach (*H*, *Kitchen&Rosenfeld3D*, *Blom3D*; first row in Fig. 10). In general, the operator *H* yields significantly larger values of ψ in comparison to the operators *Kitchen&Rosenfeld3D* and *Blom3D*; i.e., it is more difficult to separate the detections of the operator *H*. Only for images with a high noise level ($\sigma_n^2 \geq 500$) are the values of ψ for the operator *Kitchen&Rosenfeld3D* larger than those for the operator *H*. The operator *Blom3D* yields the best results of the mean curvature approach. The operators of the Gaussian curvature approach yield qualitatively similar results as the operators of the mean curvature approach (compare *K* with K^* in the second row in Fig. 10), but in absolute terms the detection performance is better (note the units of the diagrams). In comparison to these operators, the operators of the Förstner–Rohr approach (*Op3*, *Rohr3D*, and *Förstner3D*; third row in Fig. 10) yield better results. For the operators *Op3* and *Rohr3D* we obtain $\psi < 3$, while the result for the operator *Förstner3D* is worse.

The results for the tomographic images are similar to the results for the synthetic images (see Fig. 11). The operator responses of the operators of the Förstner–Rohr approach can be separated significantly better than the responses of the other operators; e.g., the operator *Rohr3D* yields in the MR as well as the CT images the lowest values for ψ. In general, the operators of the Gaussian curvature approach yield better results than the operators of the mean curvature approach. The comparison of the operators of these approaches

shows that the multiplication with the gradient magnitude improves the results (compare the operators H and K with the operators $Blom3D$ and K^*, resp.). In this investigation the results of the operator $Beaudet3D$ are worse in comparison to the operators K and K^*.

9. CONCLUSION

In this paper, we have investigated the performance of nine 3D operators for the detection of anatomical point landmarks. The mathematical definition of the operators is given in Table 1. We used quantitative measures which represent different aspects of the detection performance. We analyzed the number of corresponding points in images under elastic deformations and applied statistical measures as well as a measure of separability of the detections. Altogether each of the nine operators has been applied to 308 synthetic 3D images (of tetrahedra, ellipsoids, and hyperbolic paraboloids each with different levels of noise and deformations) and 301 tomographic (sub)images (3D ROIs of MR and CT images also with different levels of noise and deformations; the ROIs have been taken from four MR images and one CT image). Even though this work represents the currently most comprehensive study of 3D landmark detectors and significantly contributes toward a better understanding of their detection performance under noise and elastic deformation, the results must be considered preliminary until a larger-scale study is carried out using a much larger number of 3D images as well as investigating the sensitivity of acquisition parameters. The results of this investigation can be summarized as follows.

1. The operators of the Förstner–Rohr approach ($Op3$, $Rohr3D$, $Förstner3D$) are more stable w.r.t. noise than the operators of the other approaches and thus they yield a smaller number of detections. Besides that, they detect more than 70% of the landmarks in tomographic images ($P_{detect} > 0.7$). Therefore, they generally yield a larger fraction of correct detections P_{in}. Either the operator $Op3$ or the operator $Rohr3D$ yielded the largest fraction of corresponding points in deformed and noisy images. Since the largest responses of these operators are generally significantly larger than the operator responses of the other detections (performance measure ψ), a threshold could well be applied to decrease the number of false detections while preserving the correct detections. For the other operators the choice of a suitable threshold is more critical.

2. Comparing the operators of the Förstner–Rohr approach ($Op3$, $Rohr3D$, $Förstner3D$) with each other, it turns out that the operators $Op3$ and $Rohr3D$ are superior. The fraction of corresponding points in deformed and noisy images as well as the fraction of correct detections P_{in} is generally larger.

3. The multiplication with the gradient magnitude improves the results of the operators H and K as the comparison of the operators of the mean curvature (H, $Kitchen\&Rosenfeld3D$, $Blom3D$) and Gaussian curvature (K, K^*) approaches showed. The fraction of corresponding points in deformed and noisy images as well as the fraction of correct detections P_{in} of the operators $Blom3D$ and K^* are larger than those of the operators H and K, respectively. Also, the multiplication with the gradient magnitude improves the detection performance ψ.

In summary, our experimental investigation shows that the operators based on only first order partial derivatives of an image (operators $Op3$, $Rohr3D$, and $Förstner3D$) yield the

best results w.r.t. the number of corresponding points, the statistical measures, and the separability of the operator responses. Out of these operators, the operators *Op3* and *Rohr3D* show superior performance.

ACKNOWLEDGMENTS

REFERENCES

1. P. Beaudet, Rotationally invariant image operators, in *Proc. 4th Int. Joint Conf. Patt. Recog.*, *Kyoto*, *Japan*, *November 1978*, pp. 579–583.

2. W. Beil, K. Rohr, and H. S. Stiehl, Investigation of approaches for the localization of anatomical landmarks in 3D medical images, in *Proc. Computer Assisted Radiology and Surgery (CAR'97)* (H. U. Lemke, M. W. Vannier, and K. Inamura, Eds.), *Berlin*, *Germany*, *June 1997*, pp. 265–270, Elsevier, Amsterdam/Lausanne.

3. J. Blom, *Topological and Geometrical Aspects of Image Structure*, Ph.D. thesis, Utrecht University, 1992.

4. F. L. Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 567–585.

5. C. Coelho, A. Heller, J. L. Mundy, D. A. Forsyth, and A. Zisserman, An experimental evaluation of projective invariants, in *Geometric Invariance in Computer Vision* (J. L. Mundy and A. Zisserman, Eds.), pp. 87–104, MIT Press, Cambridge, MA, 1992.

6. S. Dougherty and K. W. Bowyer, Objective evaluation of edge detectors using a formally defined framework, in *Empirical Evaluation Techniques in Computer Vision* (K. W. Bowyer and P. J. Phillips, Eds.), pp. 211–234, 1998.

7. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, General intensity transformations and differential invariants, *J. Math. Imaging Vision* **4**, 1994, 171–187.

8. W. Förstner, A feature based correspondence algorithm for image matching, in *Int. Arch. Photogramm. Remote Sensing* **26**, 1986, 150–166.

9. S. Frantz, K. Rohr, and H. S. Stiehl, Multi-step differential approaches for the localization of 3D point landmarks in medical images, *J. Comput. Inform. Technol.* **6**, 1998, 435–447.

10. T. Hartkens, K. Rohr, and H. S. Stiehl, Evaluierung von Differentialoperatoren zur Detektion charakteristischer Punkte in tomographischen Bildern., in *Proc. 18. DAGM-Symposium Mustererkennung, Informatik Heidelberg, Germany, Sept. 1996* (B. Jähne, P. Geissler, and H. Haussecker, Eds.), pp. 637–644, Springer-Verlag, Berlin/Heidelberg.

11. T. Hartkens, K. Rohr, and H. S. Stiehl, Performance of 3D differential operators for the detection of anatomical point landmarks in MR and CT images, in *Proc. SPIE's International Symposium Medical Imaging, Image Processing, San Diego, CA, Feb. 1999.* (K. M. Hanson, Ed.), Vol. 3661, pp. 32–43.

12. A. Heyden and K. Rohr, Evaluation of corner extraction schemes using invariance methods, in *Proc. 13th Internat. Conf. on Pattern Recognition (ICPR'96), Vienna, Austria, Aug. 1996*, pp. 895–899, IEEE Computer Society Press, Los Alamitos, CA.

13. L. Kitchen and A. Rosenfeld, Grey level corner detection, *Pattern Recog. Lett.* **1**, 1982, 95–102.

14. K. Rohr, *Untersuchung von grauwertabhängigen Transformationen zur Ermittlung des optischen Flusses in Bildfolgen*, Diplomarbeit, Institut für Nachrichtensysteme, Universität Karlsruhe, Germany, 1987.

15. K. Rohr, Localization properties of direct corner detectors, *J. Math. Imaging Vision* **4**, 1994, 139–150.

16. K. Rohr, On 3D differential operators for detecting point landmarks, *Image Vision Comput.* **15**, 1997, 219–233.

17. K. Rohr, Extraction of 3d anatomical point landmarks based on invariance principles, *Pattern Recog.* **32**, 1999, 3–15.

18. K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. H. Kuhn, Point-based elastic registration of medical image data using approximating thin-plate splines, in *Proc. 4th Internat. Conf. Visualization in Biomedical Computing (VBC'96)* (K. H. Höhne and R. Kikinis, Eds.), Lecture Notes in Computer Science, Vol. 1131, pp. 297–306, Springer-verlag, Berlin/New York, 1996.

19. C. Schmid, R. Mohr, and C. Bauckhage, Comparing and evaluating interest points, in *Proc. Int. Conf. on Computer Vision (ICCV'98), Bombay, India, Jan. 1998*, pp. 230–235, Narosa, New Delhi/Madras, 1998.

20. J. P. Thirion and A. Gourdon, Computing the differential characteristics of isointensity surfaces, *Comput. Vision Image Understand.* **61**, 1995, 190–202.

21. J. P. Thirion, New feature points based on geometric invariants for 3d image registration, *Int. J. Comput. Vision* **18**(2), 1996, 121–137.

22. O. A. Zuniga and R. M. Haralick, Corner detection using the facet model, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1983, pp. 30–37.