Pergamon

Comput. Biol. Med. Vol. 27, No. 2, pp. 151–167, 1997

Printed in Great Britain
0010-4825/97 $17.00+0.00

PII: S0010-4825(96)00012-7

A COMPARISON OF SIMILARITY MEASURES FOR DIGITAL SUBTRACTION RADIOGRAPHY

THOMAS LEHMANN*, ABHIJIT SOVAKAR, WALTER SCHMITT[a] and RUDOLF REPGES

Institute of Medical Informatics and Biometry, Medical School, Aachen University of Technology (RWTH), D-52057 Aachen, Germany

[a] Department of Oral Surgery, Medical School, University of Bonn, Bonn, Germany

(*Received* 30 *July* 1996; *revised* 9 *December* 1996)

Abstract—Subtraction is useful in detecting small changes in sequentially acquired radiographs. Even if the imaging geometry is constant, radiographs must be registered after their digitization. To compare different algorithms for image registration and to register digital X-rays themselves, various similarity measures have been proposed. This study compares eight mathematical similarity standards using 172 radiographs acquired in different, but exactly known projection. Whenever the computation time is a critical factor, e.g. registering images using methods similar to correlation techniques, the entropy of the subtraction image's histogram function (EHDI) is found to be the best similarity standard. If not, e.g. comparative assessing different image registration techniques, the cross covariance coefficient (CCC) is appropriate. © 1997 Elsevier Science Ltd.

Quality measures Image similarity Subtraction radiography Computer assisted image processing

1. INTRODUCTION

The digital subtraction of two radiographs is the best noninvasive tool for diagnosing small changes of internal structures. For example in dental radiology, the subtraction of a series of dental films describing an identical maxillofacial region of the same patient can considerably improve not only the diagnosis of caries [1], but also the analysis of changes in bone structure caused by periodontal defects [2], by implants [3] or by guided tissue regeneration [4] regardless of whether analyses are done qualitatively or quantitatively [5–8].

Therefore, exposure conditions and projection geometry comprising X-ray source, object and sensor or film must correspond exactly. So far, this is mostly ensured using individual bite blocks [9,10] or other mechanical adjustment aids [11]. Nevertheless, X-ray films must be registered after their digitization to adjust for relative rotations and translations caused by the digitization process.

Creating digital free-hand subtraction radiographs [12], that is the X-ray subtraction technique without any mechanical devices for reproducing imaging geometry, the digital images must be adjusted for all possible movements in the projective geometry [13]. The quality of this overlay method can be judged either on the basis of subjective ROC-studies or with the aid of objective mathematical standards.

So, there are three major tasks where similarity measures are required in digital radiology:

- the determination of projection parameters (if both images are taken from the same object in different projections);
- the quality assessment of registration methods (if different algorithms for image registration are to be compared); and
- the detection of local areas with changes in internal structures (if the two images are

* Author to whom correspondence should be addressed. Tel: +49 241 80 88793; Fax: +49 241 8888 426; E-mail:lehmann@vaire.imib.rwth-aachen.de; WWW: http://www.imib.rwth-aachen.de/www/mitarb/lehmann.

Table 1. Mathematical similarity standards

Similarity measure	Abbrev.	Ref.
Cross covariance coefficient	CCC	[18]
Correlation of binary edge images	CBEI	[19]
Stochastic sign change	SSC	[20]
Sum of the absolute values of the difference image	SAVD	[21]
Standard deviation of the difference image	SDDI	[22,15]
Edges of the difference function	EDF	[23]
Standard deviation of the histogram of the difference image	SDHDI	[24]
Entropy of the histogram function of the difference image	EHDI	[25]

This table shows the abbreviations (Abbrev.) used for the mathematical similarity measures included in this study. References (Ref.) for each measure are also given.

acquired in the same projection or if they are already registrated).

In all cases, various similarity measures have been proposed, often giving only poor motivation for their usage by the authors. A systematic selection method for similarity measures is still missing and therefore still under investigation [14]. The goal of this study is to determine the power of similarity standards for dental subtraction imaging.

Section 2.1 gives a brief summary of the eight similarity measures included in this *in vitro* study. In Section 2.2 a verbal definition of image similarity in this context is translated into a mathematical definition of several comparison criteria, which are calculated on 172 test-images with different but well known projection. The image acquisition with a geometric simulation device is described in Section 2.3 and the results presented in Section 3 are discussed in Section 4 of this paper.

2. MATERIALS AND METHODS

2.1. *Mathematical similarity measures*

All selected measures (Table 1) were used in medical or dental fields, either for the comparative assessment of pairs of radiographs or to compute an automated image alignment. For their comparison, the measurements were normalized to the range between zero and one with one indicating perfect match. Therefore, their mathematical representation given below may differ with respect to their origin.

2.1.1. *Cross covariance coefficient*

In 1948, the cross correlation function was found to be the optimal detector for one-dimensional PCM-coded signals on channels with Gaussian noise [16] and is still improved in computational efficiency of biomedical signal processing [17]. In digital image processing, the cross covariance coefficient (CCC) is well known as a bias-independent measure for two-dimensional discrete data [18]:

$$CCC = \left| \frac{\sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (u(k,l) - \bar{u})(v(k,l) - \bar{v})}{\sqrt{\sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (u(k,l) - \bar{u})^2 \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (v(k,l) - \bar{v})^2}} \right| \tag{1}$$

where $\bar{u} = (1/KL)\Sigma_{k,l} u(k,l)$ denotes the mean value (bias) of the reference image $u(k,l)$ with the dimension $K \times L$ and $\bar{v} = (1/KL)\Sigma_{k,l}\, v(k,l)$ of the current image $v(k,l)$, respectively. Although the amount of computation can be reduced if (1) is implemented in the following form:

$$CCC = \left| \frac{\sum_{k,l} u(k,l)v(k,l) - KL\bar{u}\,\bar{v}}{\sqrt{\left(\sum_{k,l} u^2(k,l) - KL\bar{u}^2 \right) \left(\sum_{k,l} v^2(k,l) - KL\bar{v}^2 \right)}} \right| \tag{2}$$

correlation based measures are computationally expensive (see Section 2.2).

2.1.2. *Correlation of binary edge images*

A modified correlation method [19] was proposed in 1993. Let $B_u(k,l)$ be a binary edge image based on the image data of the reference image $u(k,l)$, e.g. thresholding the result of the Sobel-operation [18], and $B_v(k,l)$ based on the current image $v(k,l)$, respectively. A binary edge image $B(k,l) \in \{0,1\}$ is equal to one if the pixel at the image position (k,l) belongs to an edge and $B(k,l)$ is zero elsewhere. Using any edge detector as an initial step, the correlation (1) of binary edge images (CBEI):

$$CBEI = \frac{\sum_{k,l} (B_u(k,l) \cdot B_v(k,l))}{\min \left\{ \sum_{k,l} B_u(k,l), \sum_{k,l} B_v(k,l) \right\}} \tag{3}$$

can be implemented as the logical AND connection using bit-shifts instead of multiplications. The denominator in (3) is to scale CBEI to the range between zero and one. CBEI was successfully applied to the automatic compensation of relative translation of two digital dental X-rays comprising titanum implants which produce clear and sharp binary edges [19].

2.1.3. *Stochastic sign change*

In 1984, the stochastic sign change criterion (SSC) was used to correct movements of the patient in gamma-ray images [20]. Because of the system noise in gamma-ray images and in X-rays, the subtraction of two images with the same projection geometry and identical intensity parameters will not lead to a zero image, but to a normal distributed noise function with a mean value of zero. Thus, the number of sign changes along the image lines is a measure of whether images are the same (high number) or not (low number). Let $d(i)=d(k \cdot L+l)$ with $0 \le i \le K \cdot L-1$ be a linewise one-dimensional scan of the two-dimensional difference image $d(k,l)=u(k,l)-v(k,l)$ and $sc(x,y)$ a function indicating different signs (*s*ign *c*hange) in x and y, a normalized mathematical expression of this heuristic measure can be given by:

$$SSC = \frac{1}{KL-1} \sum_{i=0}^{KL-2} sc(\mathrm{d}(i), \mathrm{d}(i+1)) \text{ with } sc(x,y) := \begin{cases} 1 & \text{if } x \cdot y < 0 \\ 0 & \text{elsewhere} \end{cases} \tag{4}$$

2.1.4. *Sum of absolute values of the difference*

The use of the difference image $d(k,l)$ provides further possibilities to develop fast similarity measures. The sum of the absolute values of the difference (SAVD) was proposed in 1978 for the alignment of land-sat images [21]. Let the grey values g range from 0 to $G-1$ in both current and reference image $v(k,l)$ and $u(k,l)$, respectively. Hence, the range of the values g_{diff} of the difference image $d(k,l)$ is from $-(G-1)$ to $G-1$. Then, a normalized SAVD-measure can be written as:

$$SAVD = 1 - \frac{1}{(G-1) \cdot KL} \sum_{k,l} |\mathrm{d}(k,l)| \tag{5}$$

2.1.5. *Standard deviation of the difference image*

The standard deviation of the difference image (SDDI) was used to determine projections [15] as well as a quality measure to compare two different registration techniques for digital subtraction radiography in dental radiology [22]. The SDDI-measure can be normalized to the range between zero and one calculating the standard deviation of the worst case. Therefore, consider two chessboard-like images $u_c(k,l) \in \{0,G-1\}$ and $v_c(k,l) \in \{0,G-1\}$ with $v_c(k,l)=G-1$ if $u_c(k,l)=0$ and $v_c(k,l)=0$ if $u_c(k,l)=G-1$ for all points (k,l). The amount of the difference image $|_{dc}(k,l)|$ is $G-1$ for every pixel (k,l) and the average value $\bar{d}_c=(1/KL) \cdot \Sigma_{k,l}\ d_c(k,l)$ is zero. In this case the standard deviation $\sigma=\sqrt{1/(N-1)\Sigma_n(x_n-\bar{x})^2}$

becomes maximal:

$$\sigma_{max}=\sqrt{\frac{1}{KL-1}\sum_{k,l}(d_c(k,l)-\bar{d}_c)^2}=\sqrt{\frac{1}{KL-1}\sum_{k,l}(G-1-0)^2}\approx G-1. \tag{6}$$

Hence a normalized SDDI-function is given by:

$$SDDI=1-\frac{1}{G-1}\sqrt{\frac{1}{KL-1}\sum_{k,l}(d(k,l)-\bar{d})^2}. \tag{7}$$

2.1.6. *Edges of the difference function*

The similarity measure edges of the difference function (EDF) was published in 1994 as the base of an automatic adjustment process for digitized dental radiographs taken with standardized geometry [23]. The automatic registration procedure is able to correct translations as well as rotations resulting from misalignments during the digitization process of the dental films. Within a manually marked region of interest the edges of the difference function are computed using two specific masks [23]:

$$K_1=\begin{bmatrix}0&1&0&-1&0\\1&1&0&-1&-1\\1&1&0&-1&-1\\1&1&0&-1&-1\\0&1&0&-1&0\end{bmatrix}\text{ and }K_2=\begin{bmatrix}0&1&1&1&0\\1&1&1&1&1\\0&0&0&0&0\\-1&-1&-1&-1&-1\\0&-1&-1&-1&0\end{bmatrix} \tag{8}$$

for the convolution with the difference image $d(k,l)$. Note that the convolution with the kernel K_1 is similar to a horizontal derivation operation while K_2 represents the vertical orientation of a gradient-like operation. The results of both convolutions were combined to the EDF-measure:

$$EDF=1-\frac{1}{norm}\sum_{k,l}(|E_1(k,l)|+|E_2(k,l)|) \tag{9}$$

with $E_i=d(k,l)*K_i$ results from convolving the difference image $d(k,l)$ with the kernel K_i defined in (8). In contrast to SDDI one cannot give a worst case example for this heuristic measure. Therefore, it is impossible to determine *norm* theoretically, and the constant value *norm*=5,000,000 was chosen with respect to the image data in this study to normalize the EDF-function. However, the choice of *norm* does not influence the general order of the results.

2.1.7. *Standard deviation of the histogram of the difference image*

The last group of measures included in this study is based on the histogram function $h(g_{diff})$ of the difference image $d(k,l)=g_{diff}$ which can be defined as the relative frequency of each value $g_{diff}\in\{-(G-1),\ldots,G-1\}$:

$$h(g_{diff})=\frac{1}{KL}\sum_{k,l}\delta(d(k,l),g_{diff})\text{ with }\delta(x,y):=\begin{cases}1 & \text{if } x=y\\ 0 & \text{elsewhere}\end{cases} \tag{10}$$

The standard deviation of the histogram of the difference image (SDHDI) was proposed in 1989 to determine bone changes in periodontal defects following guided tissue regeneration [24]. Again, the chessboard-like images represent the worst case for this measure and with (6):

$$SDHDI = 1 - \frac{1}{G-1}\sqrt{\frac{1}{2G-2}\sum_{g_{diff}=-(G-1)}^{G-1}(h(g_{diff})-\bar{h})^2} \tag{11}$$

is obtained.

2.1.8. *Entropy of the histogram of the difference image*

The entropy of the histogram of the difference image (EHDI) was suggested as a quality measure in 1994 [25]. If both images u and v are identical, the entropy $H = -\Sigma h \cdot \log_2(h)$ of the histogram function $h(g_{\text{diff}})$ of the difference image $d(k,l)$ is zero. On the other hand, the entropy of an equal distributed histogram function $h_e(g_{\text{diff}}) = 1/(2G-1)$ becomes maximal, hence:

$$EHDI = 1 + \frac{1}{\log_2(2G-1)}\sum_{g_{diff}} h(g_{diff})\cdot\log_2(h(g_{diff})) \tag{12}$$

with $h(g_{\text{diff}})$ defined in (10). Note that due to the normalization in (12) the binary logarithm $\log_2$ may be replaced by any other logarithmic function.

2.2. *Definition of comparison criteria*

Two images $u(k,l)$ and $v(k,l)$ are identical if they contain the same gray values in all points (k,l). Because of system noise, two X-ray images will never be identical with respect to this definition, even if both projection geometry and X-ray dose stay constant during the acquisition of u and v. Therefore, the binary identity of two images, which can only be true or false, must be replaced by a continuous similarity of images. Beside the system noise, other factors will influence any similarity measure. For example in dental radiology, geometrical projection, X-ray dose, film speed and processing, or changes in dental regions (caries, bone destruction, tissue regeneration, etc.) results in different similarities. Whether these influences are desired or not, depends on the respective investigation.

2.2.1. *Resolution*

To be used in digital free-hand subtraction radiography [12,13], a similarity measure should be independent of the discrete image resolution $K \times L$. Only then, the similarities are absolutely comparable regardless of the number of pixels created by the X-ray sensor or the CCD-camera used to digitize the radiographs. In other words, a measure is truly independent of the image resolution if it can provide the same result when the sizes of the images are changed without changing their contents. Hence:

$$SIM(u(k,l),v(k,l)) \stackrel{!}{=} SIM(u(p,q),v(p,q)) \tag{13}$$

with $\stackrel{!}{=}$ expressing the requirement of equality and $p=0,\ldots,P-1$; $q=0,\ldots,Q-1$; $K \neq P$ and $L \neq Q$, respectively. The function SIM(u,v) in (13) denotes any of the similarity measures defined in Section 2.1.

2.2.2. *Intensity*

A second criterion is given by the standard's independence of the X-ray dose and the film speed, emulsion and processing. If direct digital sensors are used, the X-ray dose is the significant parameter for the image intensity. In a first approach, any measure must therefore be independent of the pixel's ground value (bias):

$$SIM(u(k,l),v(k,l)) \stackrel{!}{=} SIM(u(k,l)+c_1,v(k,l)+c_2) \tag{14}$$

where $c_i \in \text{IR}$ is any constant value.

2.2.3. *Linearity*

Linear movements of the image aperture scanning the object should result in a linear progression of the similarity curve. In other words, if two radiographs cover a larger region,

their computed similarity should be higher; if both images are identical, their similarity should be one and the similarity of two completely different radiographs (no overlapping region) should be zero. Based on series $v_i(k,l)$ of dental X-rays with exactly known imaging geometry, the suitability of the measures SIM(u,v) introduced in Section 2.1 can be evaluated by the variation about the ideal progression IDEAL(u,v):

$$\sigma^2 = \frac{1}{N} \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} (SIM(u,v_i) - IDEAL(u,v_i))^2 \tag{15}$$

where N denotes the odd number of images $v_i(k,l)$ contained in each series. The middle image $v_0(k,l)$ was chosen as the reference image $u(k,l)$ in all sequences acquired for this study (see Section 2.3).

2.2.4. *Runtime*

Using similarity standards for image registration methods related to the cross correlation, the measures have to be calculated for every possible matching position. Therefore, the computation of the measures must be easy and fast. In this study, the costs of computation are determined by the number and kind of operations required to calculate the similarity multiplicated with the exactly measured average computing time needed for these operations.

2.3. *Test-image acquisition*

In a recently published study [14] the appropriate statistic for identifying changes in pairs of dental radiographs could not be selected based on simulations. The determination of (15) yields related problems. Because of difficulties simulating the influence of different image contents in the non-overlapping parts of the radiographs to be compared, the evaluation of (15) is founded on *in vitro* X-rays.

Therefore, a mechanical device was constructed to guarantee the geometrical fixation of tube, object and sensor (Fig. 1). Displacements of the sensor can be adjusted with a precision of 1/100 mm using micrometer screws; the scales on the revolving axes allow the

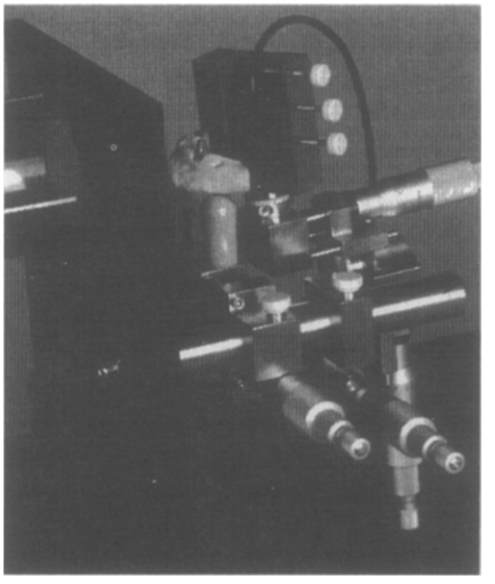

Fig. 1. Device for image acquisition. This figure shows the mechanical device constructed for the *in vitro* study. On the upper left a part of the mounted tube is visible. The X-rays pass the phantom and were converted into a digital image with the aid of the CCD-sensor, shown in the upper right. The dose-, rotation-, scaling- and translation-sequence (Figs 3, 5, 7 and 9, respectively) are produced using the Sens-A-Ray system (REGAM Medical Systems, Sundsvall, Sweden). The micrometer screws allow the adjustment of translations with a precision of 1/100 mm. The scales on the revolving axes can be adjusted in steps of 1°. Note that the phantom shown in this figure differs from the *in vitro* object used in this study (Fig. 2).

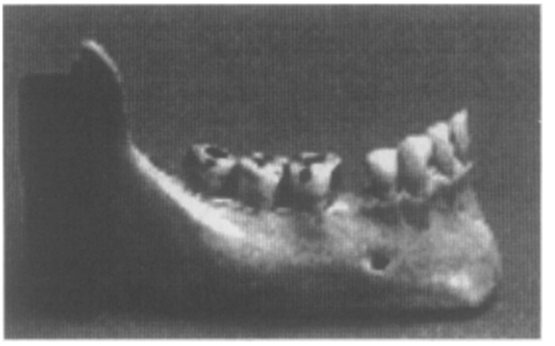

Fig. 2. *In vitro* object. This part of a human mandible was used to create the sequences shown in Figs 3, 5, 7 and 9. Therefore, the dry jawbone was mounted on the adjustment device (Fig. 1).

adjustment of rotations with a precision of 1°. For this study, the simulation device was combined with a direct digital image acquisition system (Sens-A-Ray, REGAM Medical Systems, Sundsvall, Sweden). A dry mandible complete with teeth and fillings (Fig. 2) was mounted on the device and used as an *in vitro* object.

A total of four different series of X-rays has been acquired. The dose sequence (Fig. 3) was produced to analyse the measure's dependence to the X-ray dose, keeping the system's

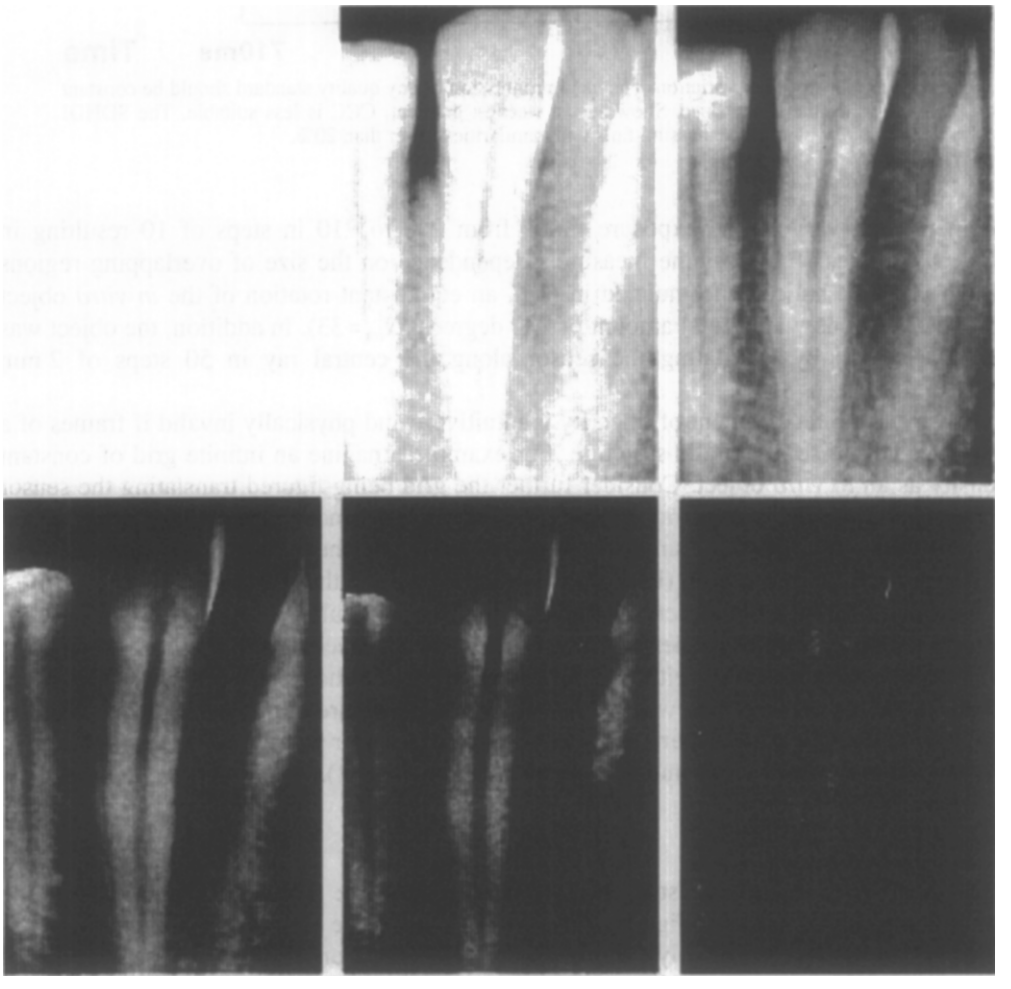

Fig. 3. Dose sequence. The projection geometry remained unchanged while the X-ray dose was varied. Starting with 250 (scale of the timer unit, REGAM Medical Systems, Sundsvall, Sweden) the exposure time was raised up to 710 in steps of 10 resulting in $N_{dose}=47$ radiographs. The images v_i, $i=-23,\ -14,\ -5,\ 5,\ 14$ and 23 are shown in the upper left, upper middle, upper right, lower left, lower middle and lower right, respectively.

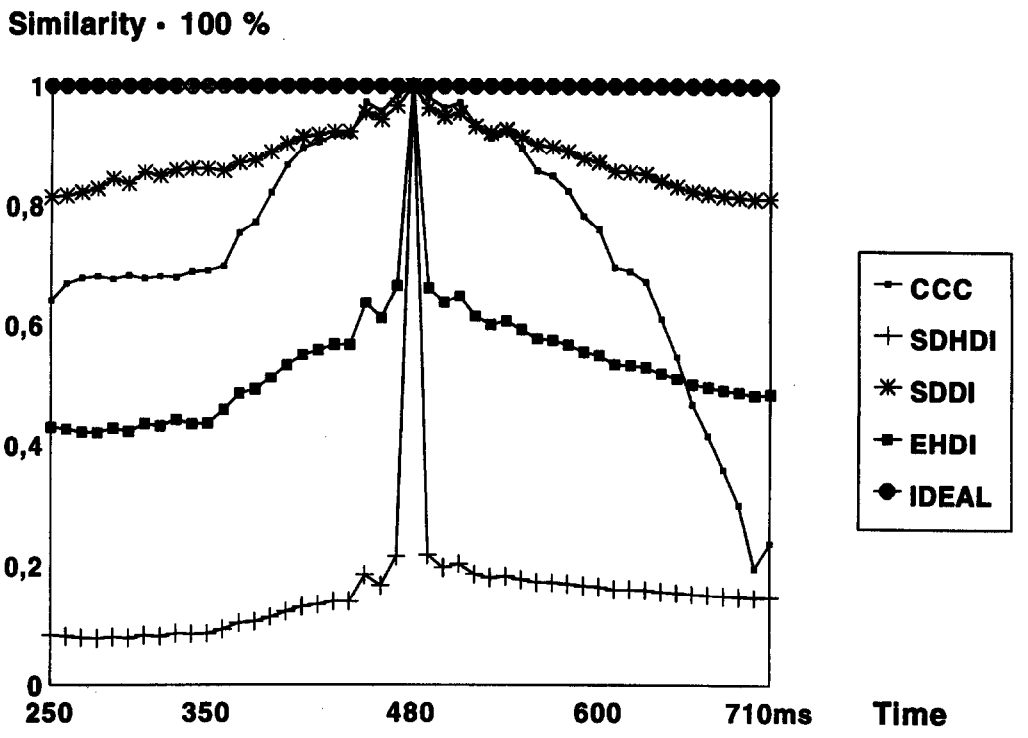

Fig. 4. Results of dose variation. The ideal progression of any quality standard should be constant varying only the X-ray dose. Showing the steepest gradient, CCC is less suitable. The SDHDI measure fails with similarities lower than 20%.

geometry and varying the exposure time* from 250 to 710 in steps of 10 resulting in $N_{dose}=47$ frames. To assess the measure's dependence on the size of overlapping regions which is mathematically formulated in (15), an equidistant rotation of the *in vitro* object (Fig. 5) was done with a total amount of 180 degrees ($N_{rot}=33$). In addition, the object was scaled (Fig. 7) by translating the sensor along the central ray in 50 steps of 2 mm ($N_{scale}=51$).

However, the requirement of linearity is intuitively and physically invalid if frames of a sequence comprise a periodic structure. For example, imagine an infinite grid of constant density as an *in vitro* object. Consider further the grid being figured translating the sensor perpendicular to the central ray but along the grid. This movement should result in the periodic behaviour of every similarity function. Note that there is a pseudo-periodicity in the dentomaxillofacial region (Fig. 2) for movements along the jaw, too.

Nevertheless, the requirement of linearity may be particularly desired in such pseudo-periodical cases. For example, if optimized correlation techniques are used to register images. In most of those algorithms a similarity only along the steepest path of the gradient is computed. Such processes will stop if continuous teeth are aligned. The fourth set was realized to evaluate pseudo-periodicity translating the sensor parallel to the mandible (Fig. 9) with a total amount of 40 mm in 40 equal steps ($N_{trans}=41$).

3. RESULTS

The results of this *in vitro* study are summarized in Table 5 relating the columns to the equations given in the text. Figures 4, 6, 8 and 10 show the results of the four test-series (Figs 3, 5, 7 and 9, respectively). Since CBEI, SSC, SAVD and EDF do not satisfy (13) or (14) (see below) they are not plotted. IDEAL denotes the progression of an ideal similarity measure.

* Scale of the timer unit, REGAM Medical Systems, Sundsvall, Sweden.

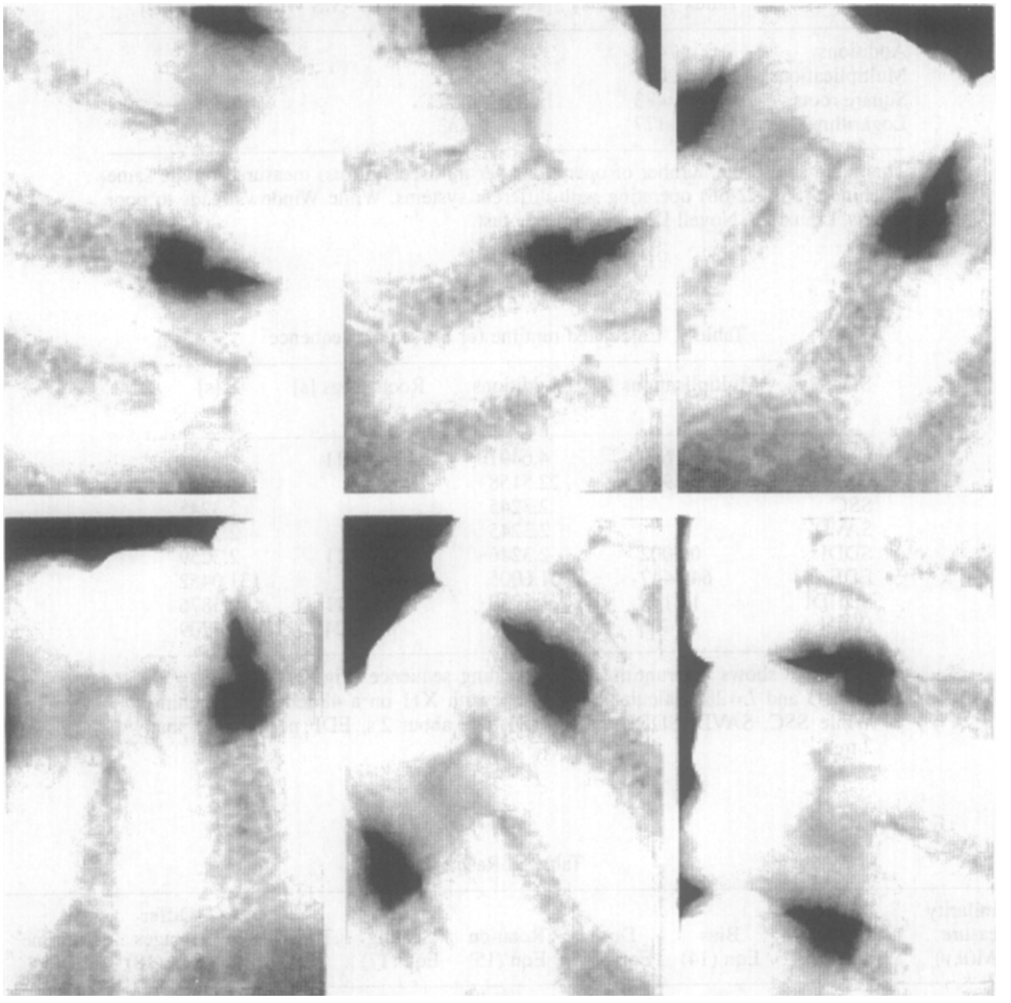

Fig. 5. Rotation sequence. Acquiring the rotation series, the *in vitro* object (Fig. 2) was fixed in front of the tube. Keeping the position of the sensor constant, the X-ray source with the *in vitro* object was 180° rotated in equidistant steps producing $N_{rot}=33$ radiographs. The images v_i, $i=-16$, -10, -3, 3, 10 and 16 are shown in the upper left, upper middle, upper right, lower left, lower middle and lower right with a relative rotation to the (not shown) reference image $u=v_0$ of $-90°$, $-52°$, $-17°$, 17°, 52° and 90°, respectively.

Table 2. Required operations

SIM(u,v)	Multiplications	Additions	Roots	Logarithms
CCC	$(N-1)\cdot 3KL$	$(N-1)\cdot 4KL$	$N-1$	
CBEI	$N\cdot 18KL$	$N\cdot 19KL$		
SSC		$(N-1)\cdot 2KL$		
SAVD		$(N-1)\cdot 2KL$		
SDDI	$N\cdot 7$	$(N-1)\cdot 2KL$	N	
EDF	$(N-1)\cdot 50KL$	$(N-1)\cdot 53KL$		
SDHDI	$(N-1)\ (KL+2G_{diff})$	$(N-1)(KL+G_{diff})$	N	
EHDI	$(N-1)\ (KL+4G_{diff})$	$(N-1)(KL+2G_{diff})$		$(N-1)G_{diff}$

This table shows the number of multiplications, additions, square-roots and logarithms required to compute the similarity measures SIM (u,v). The given orders are approximations resulting from optimized implementation and so, they may differ from the equations defined in the text. $K\times L$ denotes the discrete image dimensions, $G_{diff}=2G-1$ the range of the difference image and N the number of images to be compared.

Table 3. Measured runtimes

Operations	Linux 1.1.62 [ms]	Novell Dos 7 [ms]	MS Windows 3.1 [ms]
Additions	2439	2632	552
Multiplications	2041	2273	534
Square roots	483	457	196
Logarithms	177	83	76

This table shows the number of operations per millisecond (ms) measured on the same machine (486DX2-66) operating with different systems. While Windows leads to poor results, Linux and Novell Dos are similarly fast.

Table 4. Calculated runtime for the scaling sequence

SIM(u,v)	Multiplications [s]	Additions [s]	Roots/Logs [s]	Σ [s]
CCC	4.1669	4.6491	0.00011	8.8161
CBEI	25.4908	22.5158		48.0065
SSC		2.3245		2.3245
SAVD		2.3245		2.3245
SDDI	0.0002	2.3246	0.00011	2.3259
EDF	64.4447	61.6005		131.0452
SDHDI	1.4145	1.1730	0.00011	2.5876
EHDI	1.4400	1.1837	0.14724	2.7709

The table shows the runtime for the scaling sequence (Fig. 7) with $N_{scale}=51$, $K=193$ and $L=288$ calculated for Linux with X11 on a 486DX2-66 machine. While SSC, SAVD, SDDI and SDHDI take about 2 s, EDF needs more than 2 min.

Table 5. Results

Similarity measure SIM(u,v)	Resolution Eqn (14)	Bias Eqn (14)	Dose Eqn (15)	Rotation Eqn (15)	Scaling Eqn (17)	Translation Eqn (15)	Differ. images Eqn (18)	Runtime Table 4
CCC	Yes	Yes	0.107	0.085	0.06	0.083	0.143	8.82
CBEI	No	Yes	0.639	0.135	0.16	0.191	0.139	48.0
SSC	Yes	No	0.986	0.225	0.14	0.258	0.044	2.32
SAVD	Yes	No	0.883	0.220	0.06	0.124	0.759	2.32
SDDI	Yes	Yes	0.017	0.154	0.08	0.092	0.690	2.33
EDF	No	Yes	0.195	0.080	0.26	0.065	0.512	131.0
SDHDI	Yes	Yes	0.726	0.208	0.08	0.227	0.064	2.59
EHDI	Yes	Yes	0.226	0.065	0.08	0.089	0.264	2.77
IDEAL	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0

The table summarizes the results of this study. Since the measures CBEI and EDF do not satisfy (13) and SSC and SAVD do not satisfy (14), they are not plotted in Figs 4, 6, 8 and 10.

Table 6. Resulting scores

Similarity measure	Dose	Rotation	Scaling	Translation	Different images	Score Σ_1	Runtime	Score Σ_2
CCC	0.148	0.407	0.750	0.366	0.207	1.878	1.0	2.878
SDDI	0.023	0.740	1.0	0.406	1.0	3.169	0.264	3.433
SDHDI	1.0	1.0	1.0	1.0	0.093	4.093	0.294	4.387
EHDI	0.311	0.310	1.0	0.394	0.383	2.397	0.314	2.711

Since CBEI, SSC, SAVD and EDF do not satisfy (13) or (14), they are not included in the calculation of the resulting score. All scores have been taken from Table 5 and normalized to the highest number in each column. If the runtime is unimportant CCC should be preferred (score Σ_1) else EHDI is shown to be the best similarity measure (score Σ_2).

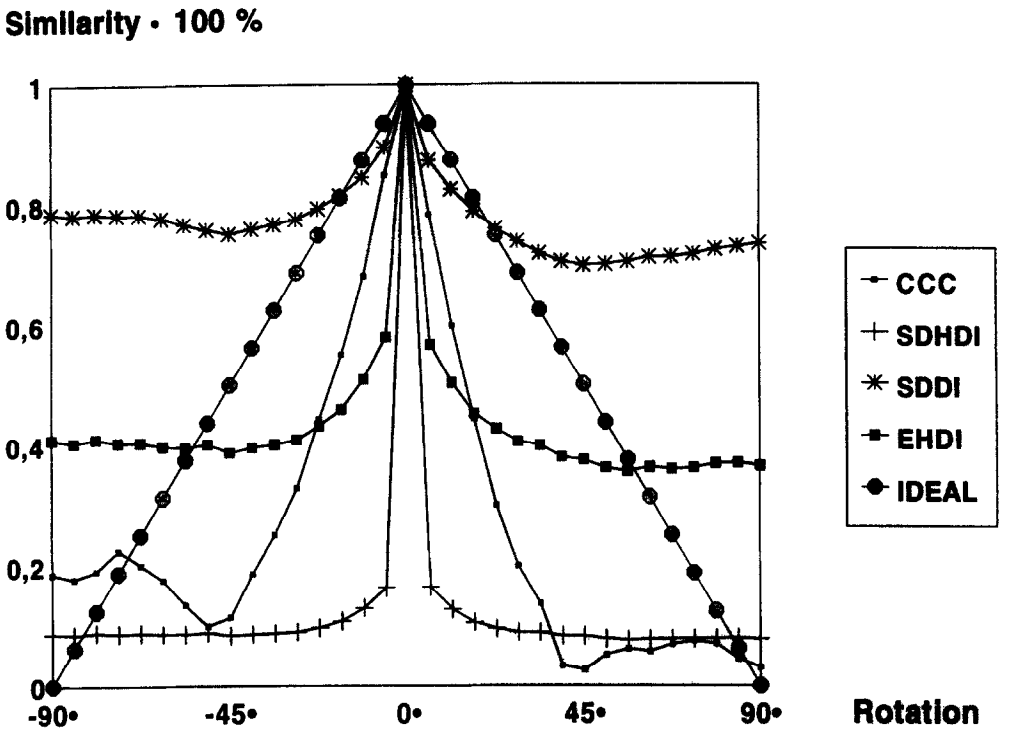

Fig. 6. Results of rotation. The curves of the quality measures were computed with the aid of the rotation sequence (Fig. 5). While the SDDI-curve is ascending in the interval from 45° to 90°, the EHDI measure shows a nearly monotonous progession. CCC is not monotonic and SDHDI is almost always constant.

3.1. *Resolution*

CBEI and EDF include an edge detection step which is computed using convolution masks of fixed dimensions. Therefore, both measures depend on the discrete image resolution and do not satisfy (13) defined in Section 2.2.

3.2. *Intensity*

The independence on the image bias which approximates the independence on the X-ray dose is demanded in (14). Let $d_1(k,l)=u(k,l)-v_1(k,l)$ and $d_2(k,l)=u(k,l)-v_2(k,l)$ be the differences between the reference u and the consecutive images v_i with $v_2(k,l)=v_1(k,l)+c$. Then (5) yields:

$$SAVD_1 - SAVD_2 = C\cdot\left(\sum_{k,l}|d_2| - \sum_{k,l}|d_1|\right) = C\cdot\sum_{k,l}(|d_2|-|d_1|) = C\cdot\sum_{k,l}|c| \neq 0 \qquad (16)$$

where $C=1/(KL(G-1))$ and $c\in\mathbb{R}^+$ is some constant value. The same results can be shown referring to (4). So the requirement expressed in (14) holds for neither SAVD nor SSC. The evaluation of the dose sequence (Fig. 4) experimentally confirms this mathematical investigation. As expected, SSC and SAVD resulted in the worst case (Table 5).

3.3. *Linearity*

Linear movements without pseudo-periodical structures are obtained from the rotation and the scaling sequence. While the rotation sequence was analysed with (15) resulting CCC and EHDI best (Table 5), an *a priori* ideal curve for the scaling sequence (Fig. 8) could not be given. Because the endpoint, i.e. the particular scale where the similarity should be zero, is not known, the slope of the ideal line is not determinable. Therefore, the required linearity is extenuated to monotonicity quantitatively evaluated counting the number of sign changes in the gradient of the similarity measure:

$$SIM_{scale} = \frac{1}{N_{scale}-1}\left(\sum_{i=-(N_{scale}-1)/2}^{-1} sc(1,W(i)) + \sum_{i=1}^{(N_{scale}-1)/2} sc(-1,W(i-1))\right) \qquad (17)$$

with $W(i)=\mathrm{SIM}(v_0,v_{i+1})-\mathrm{SIM}(v_0,v_i)$ and $sc(x,y)$ defined in (4). As expected, the similarity measures based on a convolution with a kernel of fixed size: EDF and CBEI obtained only poor results. Since the scaling sequence was acquired with a constant X-ray intensity moving the sensor away from the object and keeping the position of the X-ray source and the object (Fig. 7), the SSC measure, which depends on the image bias, is also bad in this case.

The results of an equidistant translation of the sensor perpendicular to the central beam of the X-ray source are shown in Fig. 10. The active area's width of the Sens-A-Ray sensor is 17.2 mm [26]. Therefore, the variance defined in (15) is calculated including 35 images ($(N-1)/2=17$). The best results are obtained with EDF and EHDI. CCC shows the most periodical behaviour.

The similarity of completely different radiographs should ideally be zero. In this study the similarities of radiographs of different dental regions are averaged using the first three and the last three images of the translation sequence with SSC and SDHDI coming off well

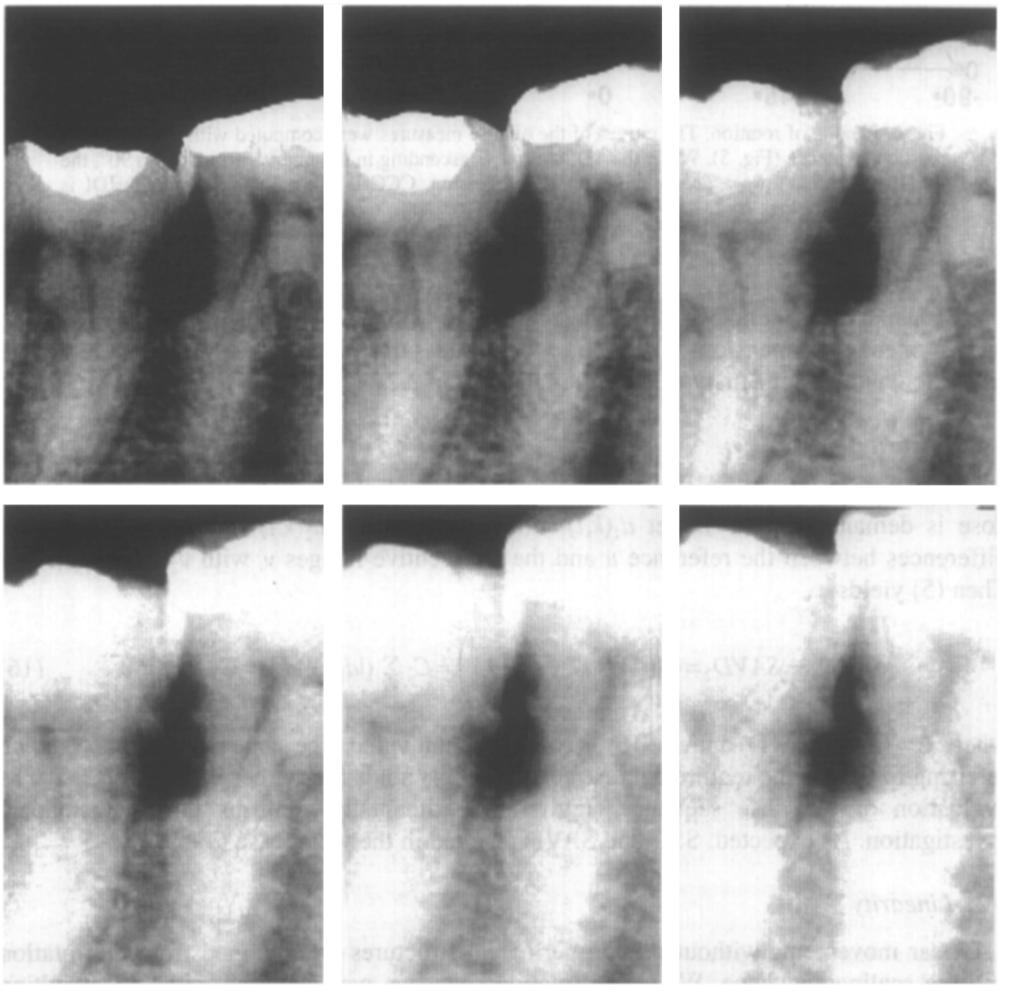

Fig. 7. Scaling sequence. The scaling sequence with $N_{scale}=51$ radiographs was produced moving the sensor in steps of equal distance parallel to the central ray of the X-ray projection with a total amount of 10 cm. The images v_i, $i=-25,\ -15,\ -5,\ 5,\ 15$ and 25 are shown in the upper left, upper middle, upper right, lower left, lower middle and lower right, respectively. Note that the radiation dose remained unchanged. Therefore, in the frames v_{-25} and v_{-15} saturation effects of the CCD-element are clearly visible while the sequence's last radiograph v_{25} seems to be underexposed.

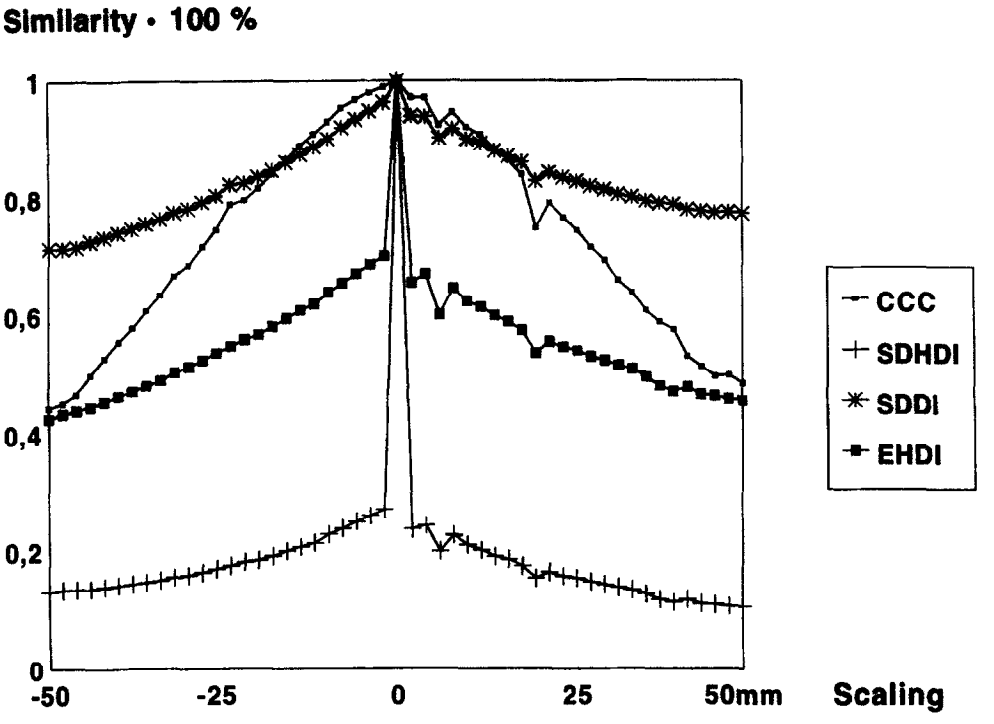

Fig. 8. Results of scaling. The scaling sequence (Fig. 7) leads to this similarity curves. The SDHDI, SDDI and EHDI quality measures obtain similar results. The steepest progression is produced using CCC. A misarrangement of the sensor acquiring the pictures v_3 and v_{10} produces artefacts visible in all curves.

(Table 5):

$$SIM_{diff} = \frac{1}{6}\left(\sum_{\substack{i=-(N_{trans}-1)/2 \\ |i|>17}}^{(N_{trans}-1)/2} SIM(v_0, v_i)\right). \tag{18}$$

3.4. *Runtime*

All measures are implemented on a 486DX2-66 PC operating with Linux, Novell Dos and Windows. Table 2 shows the numbers of multiplications, additions, square roots and logarithms required to compute each measure referring to its efficient implementation, not to its mathematical definition presented in Section 2.1. On the other hand, the numbers of operations per millisecond were exactly determined (Table 3). The resulting runtimes for the scaling sequence are presented in Table 4. Two convolutions with kernels of 5×5 pixels are neccessary to compute EDF. Therefore, EDF takes 50 times of the averaged runtime (2.5 s) of the fast measures: SSC, SAVD, SDDI, SDHDI and EHDI. CBEI is 20 times (here the convolution masks have the dimension 3×3) and CCC is after all 3.5 times slower than the others.

4. DISCUSSION

The results summarized in Table 5 were normalized and linearly combined resulting in the scores shown in Table 6. Only four of the eight measures compared in this study are independent of the discrete image resolution and the pixels mean value (bias). Since CBEI, SSC, SAVD and EDF do not satisfy (13) or (14) they are not scored.

SDDI, SDHDI and EHDI are all based on the difference image, which leads to similar good results. SDDI does not sufficiently distinguish between different images (similarities between 70% and 80% in Figs 6, 7 and 8). Although in this case SDHDI is the best measure, it is the worst in all other criteria. This is caused by normalization, necessary to handle the

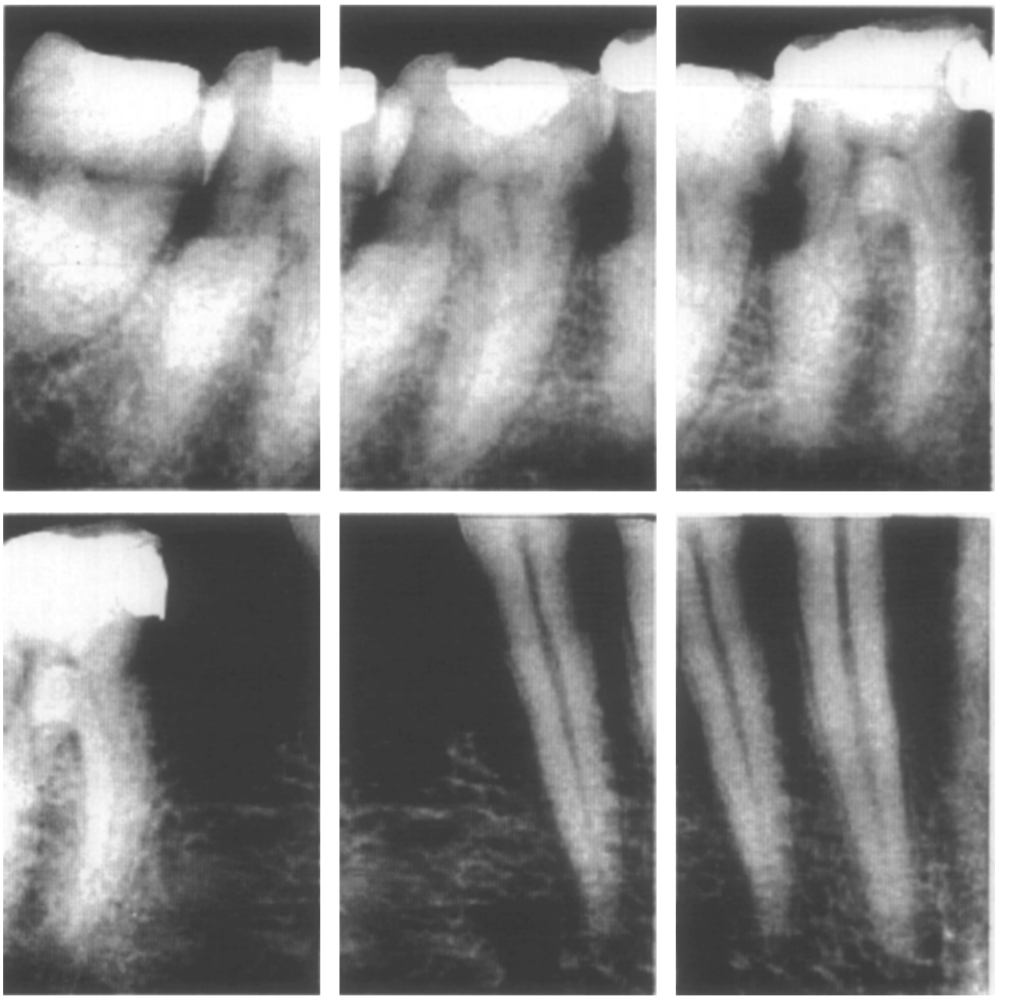

Fig. 9. Translation sequence. The translation sequence consists of $N_{trans}=41$ images produced by moving the sensor along the jaw (Fig. 2) and perpendicular to the central ray of the X-ray tube in equal steps of 1 mm. The images v_i, $i=-20, -12, -4, 4, 12$ and 20 of the translation sequence are shown in the upper left, upper middle, upper right, lower left, lower middle and lower right, respectively. Note that there is no overlapping region between the (not Shown) reference image $u=v_0$ and the first image v_{-20} or the last image v_{20}.

worst case for the SDHDI measure: two complementary chessboard-like images with d_c from (6) leading to $h_c(G-1)=h_c(1-G)=1/2$, and $h_c(g_{diff})=0$ elsewhere. In contrast, the normalization of EHDI is the constant histogram function of the difference image $h_c(g_{diff})=1/(2G-1)$. So the necessary normalization of the standard deviation to the range between zero and one suppresses the similarity measure's dynamics (see Figs 6 and 9). CCC shows supreme dynamics in all series but also most non-monotonic behaviour on pseudo-periodical data (Fig. 10).

There are utilizations of similarity standards in digital radiology where the computation time is secondary. Therefore, score Σ_1 in Table 6 was generated ignoring the runtime. In this case, CCC is superior to all the other measures. Note that this result stays constant regardless of whether pseudo-periodical structures are considered (column "Translation" in Table 6) or not.

Taking into account the computation time EHDI was found to be the best similarity measure (score Σ_2 in Table 6). The rapidness of EHDI is founded in the reduction of dimensions. The two-dimensional image analysis is reduced to a one-dimensional histogram interpretation. It should be pointed out that this result is independent of the consideration of pseudo-periodical structures (column "Translation" in Table 6).

As a result of this study, image similarity measurement was found to be another domain

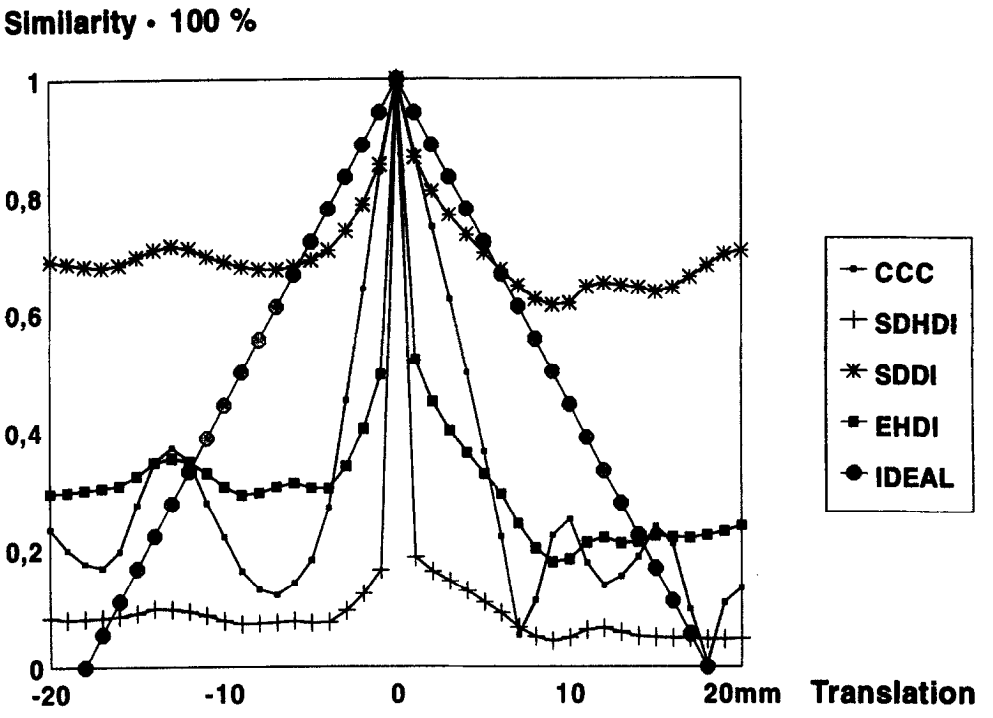

Fig. 10. Results of translation. This similarity curves are computed using the middle image of the translation sequence (Fig. 9) as reference. Because of the pseudo-periodical structure, the CCC-curve shows a lot of peaks. Depending on the kind of investigation a monotonous progression may be desired instead.

where the analysis of the histogram's entropy yields convenience. The entropy's classical area in computer vision, graphics and image processing is the histogram based grey-level thresholding [27] where entropy analysis has been proven to be reasonably good [28].

Finally in future, in the digital free-hand subtraction radiography project, the cross covariance coefficient CCC will be used as a quality standard to evaluate different modules for automatical alignment of X-ray images, while the entropy of the subtraction image's histogram EHDI will be used whenever a correlation step has to be realized registering radiographs.

5. SUMMARY

This study compares eight mathematical similarity standards which have been used in medical image processing. The objective is to evaluate the power of these measures for the developement of medical image registration techniques. Based on its verbal definition, image similarity was formulated mathematically resulting in quality criteria. Those criteria were evaluated on various *in vitro* data to guarantee objective results. In order to acquire radiographs in different, but exactly known projections, a mechanical device with micrometer screws was constructed. Four X-ray sequences containing 172 frames were captured varying the projection geometry and X-ray dose.

Whenever computation time is critical, e.g. registering images using methods similar to correlation techniques, the entropy of the subtraction image's histogram function (EHDI) was found to be the best similarity standard. If not, e.g. comparative assessing different image registration techniques, the cross covariance coefficient (CCC) is appropriate.

Acknowledgements—The project "Free-hand subtraction radiography" is funded by the German Research Community DFG (Re 427/5-1). The authors would like to thank Ingrid Bergener, Institute of Medical Informatics and Biometry, Aachen University of Technology, Germany, for the helpful discussions and Renate Meyer, School of Mathematical and Information Sciences, University of Auckland, New Zealand, for her assistance in translation.

REFERENCES

1. H.G. Gröndahl, K. Gröndahl, T. Okano and R.L. Webber, Statistical contrast enhancement of subtraction images for radiographic caries diagnosis, *Oral Surg.*, **53**, 219–223 (1982).
2. H.G. Gröndahl and K. Gröndahl, Subtraction radiography for the diagnosis of periodontal bone lesions, *Oral Surg.*, **55**, 208–213 (1983).
3. M.S. Reddy, M.K. Jeffcoat and R.C. Richardson, Assessment of adjunctive flurbiprofen therapy in root-form implant healing with digital subtraction radiography, *J. Oral Implantol.*, **16**, 272–276 (1990).
4. A. Wenzel, K. Warrer and T. Karring, Digital subtraction radiography in assessing bone changes in periodontal defects following guided tissue regeneration, *J. Clin. Periodontol.*, **19**, 208–213 (1992).
5. L.F. Ortman, R. Dunford, K. McHenry and E. Hausmann, Subtraction radiography and computer assisted densitometric analyses of standardized radiographs: A comparison study with ^{125}I absorptiometry, *J. Periodont. Res.*, **20**, 644–651 (1985).
6. M.H. Vos, P.T.M. Janssen, J. van Aken and R.M. Heethaar, Quantitative measurement of periodontal bone changes by digital subtraction, *J. Periodont. Res.*, **21**, 583–591 (1986).
7. D. Brägger, L. Pasquali, H. Rylander, D. Carnes and K.S. Kornman, Computer-assisted densitometric image analysis in periodontal radiography: A methodological study, *J. Clin. Periodontol.*, **15**, 27–37 (1988).
8. A. Mol and P.F. van der Stelt, Application of digital image analysis in dental radiography for the description of periapical bone lesions: a preliminary study, *IEEE Trans. Biomed. Engng*, **38**, 357–359 (1991).
9. A.S. Duinkerke, A.C. van de Poel, F.P. van der Linden, W.H. Doesburg and W.A. Lemmens, Evaluation of a technique for standardized periapical radiographs, *Oral Surg.*, **44**, 646–651 (1977).
10. D.J. Rudolph and S.C. White, Film-holding instruments for intraoral subtraction radiography, *Oral Surg. Oral Med. Oral Pathol.*, **65**, 767–772 (1988).
11. M.K. Jeffcoat, M.S. Reddy, R.L. Webber, R.C. Williams and U.E. Ruttimann, Extraoral control of geometry for digital subtraction radiography, *J. Periodont. Res.*, **22**, 396–402 (1987).
12. T. Lehmann and W. Schmitt, Determination of bone density in dental implantology by automatic subtraction of digital free-hand radiographs taken intraorally. In: D.S. You, T.W. Park and S.R. Lee, eds. Proceedings of the 10th International Conference of Dento-Maxillo-Facial Radiology, Seoul, Korea, pp. 188–192 (1994).
13. T. Lehmann and W. Schmitt, Intraoral imaging by digital radiography and digital image processing, *Dentomaxillofac. Radiol.*, **23**, 61 (1994).
14. M.E. Cohen and W.C. Roddy, A comparison of three statistics for detecting differences in digitized dental radiographs: a simulation study, *Dentomaxillofac. Radiol.*, **24**, 179–184 (1995).
15. P.F. van der Stelt, U.E. Ruttimann and R.L. Webber, Determination of projections for subtraction radiography based on image similarity measurements, *Dentomaxillofac. Radiol.*, **18**, 113–117 (1989).
16. B.M. Oliver, R.J. Pierce and C.E. Shannon, The philosophy of PCM, *Proc. IRE*, **36**, 1324–1332 (1948).
17. R.D. Throne, Detecting ventricular fibrillation using efficient techniques for computing a normalized autocorrelation, *Comput. Biol. Med.*, **23**, 317–325 (1993).
18. W.K. Pratt, *Digital Image Processing*. John Wiley, New York (1978).
19. W. Schmitt and T. Lehmann, Digitale Radiographie und digitale Bildverarbeitung in der implantologischen Diagnostik, *Z. Zahnärztl. Implantol.*, **9**, 284–287 (1993).
20. A. Venot, J.F. Lebruchec and J.C. Roucayrol, A new class of similarity measures for robust image registration, *CVGIP*, **28**, 176–184 (1984).
21. M. Svedlow, C.D. McGillem and P.E. Anuta, Image registration: similarity measure and preprocessing method comparisons, *IEEE AES*, **14**, 141–149 (1978).
22. S.M. Dunn, P.F. van der Stelt, K. Fenesy and S. Shah, A comparison of two registration techniques for digital subtraction radiography, *Dentomaxillofac. Radiol.*, **22**, 77–80 (1993).
23. J. Samarabandu, K.M. Allan, E. Hausmann and R. Acharya, Algorithm for the automated alignment of radiographs for image subtraction, *Oral Surg. Oral Med. Oral Pathol.*, **77**, 74–79 (1994).
24. A. Wenzel, Effect of manual compared with reference point superimposition on image quality in digital subtraction radiography, *Dentomaxillofac. Radiol.*, **18**, 145–150 (1989).
25. T. Lehmann, W. Schmitt, R. Repges and A. Sovakar, Mathematical quality standards for the digital free-hand subtraction radiography, *Dentomaxillofac. Radiol.*, **24**, 98 (1995).
26. U. Welander, P. Nelvig, G. Tronje, W.D. McDavid, S.B. Dove, A.C. Mörner and T. Cederlund, Basic technical properties of a system for direct acquisition of digital intraoral radiographs, *Oral Surg. Oral Med. Oral Pathol.*, **75**, 506–516 (1993).
27. J.N. Kapur, P.K. Sahoo and A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *CVGIP*, **29**, 273–285 (1985).
28. P.K. Sahoo, S. Soltani, A.K.C. Wong and Y.C. Chen, A survey of thresholding techniques, *CVGIP*, **41**, 233–260 (1988).

About the Author—THOMAS LEHMANN was born in Bonn, Germany in 1966. He received the Masters degree in Electrical Engineering from the Aachen University of Technology in 1992. In 1992 he was a scientist at the Institute for Measurement Technology, Aachen University of Technology. Since 1992 he has been with the Institute of Medical Informatics and Biometry, Aachen University of Technology, and has been the Head of the Department of Medical Image Processing since 1995.

In 1993 he was given the DAGM-Preis '93. The award from the German Association for Pattern Recognition (DAGM) was given for his work on automatical strabometry using Hough-transform and covariance filtering. Since 1996 he has been the Editor and Chair of the Aachen Workshop on Medical Image Processing. In 1997 his textbook on image processing in medicine will be published by Springer-Verlag, Germany.

He is member of the Institute of Electrical and Electronics Engineers (IEEE), the International Society for Optical Engineering (SPIE), and the International Association of Dento-Maxillo-Facial Radiology (IADMFR). His research interests are in medical image processing and its applications

to computer aided diagnosis.

About the Author—ABHIJIT SOVAKAR was born in Calcutta, India in 1965. Since 1990 he has studied computer science at the Aachen University of Technology. In 1996 he finished his Masters thesis on contour tracking of vocal folds in videostroboscopic image sequences for which he has been nominated for the Sun Microsystems Student Award.

About the Author—WALTER SCHMITT was born in 1950 in Trier, Germany. In 1975 he received the Masters degree in Electrical Engineering at the Aachen University of Technology. In 1982 he earned the Dental Doctor at the Research Institute for Experimental Surgery, Davos, Switzerland. He became a Professor for Oral Surgery at the Aachen University of Technology in 1992.

From 1975 until 1977 he was with the Institute of Mechanical Construction Design, Aachen University of Technology. In 1983 he joined the Clinic for Orthodontics, University of Freiburg, Germany. From 1984 to 1995 he was with the Clinic for Oral Surgery, Aachen University of Technology. Since 1995 he has been with the Clinic for Oral Surgery at the University of Bonn. His research interests are in dental implantology and digital radiology where he published about 45 papers.

About the Author—RUDOLF REPGES was born in Wesel, Germany in 1927. He obtained his education in Mathematics and Physics in Aachen and Gießen, Germany while his education in Medicine was obtained in Cologne, Freiburg and Gießen, Germany. He was assistant professor in Physiology and Biochemistry at the University of Gießen, Germany, where he received the Venia Legende for Biomathematics in 1968.

From 1971 until 1990 he was full professor of Medical Statistics and Documentation at the Aachen University of Technology, Aachen, Germany. From 1990 until 1995 he was full professor for Medical Informatics and Biometry at the Aachen University of Technology, Aachen, Germany.

He was president of the Biometric Society, German region, and in the board of the German Society of Medical Informatics, Biometry and Epidemiology. He is the author of some textbooks and about 100 scientific papers.