Gray Hausdorff distance measure for medical image comparison in dermatology: Evaluation of treatment effectiveness by image similarity

Panagiota Spyridonos1,*, Georgios Gaitanis2, Ioannis D. Bassukas2 and Margaret Tzaphlidou1

1Department of Medical Physics, University of Ioannina Medical School, University Campus, Ioannina, Greece and 2Department of Skin and Venereal Diseases, University of Ioannina Medical School, University Campus, Ioannina, Greece

Abstract: Introduction: In clinical dermatology, the stabilization of the overall skin condition can be in many cases the earliest qualitative measure of the effectiveness of the therapeutic intervention. Subjective image comparisons, that offer empirical ‘qualitative’ judgments of degrees of image similarities, are traditionally employed by the involved physicians.

Objectives: To quantify, by means of an image similarity metric, the degree of stabilization of an expanding skin disease, and to identify the situation of ‘no further change’ of the skin condition of the patient, providing thus the physician with an early, objective measure of the efficacy of the used therapy.

Methods: For treatment assessment, a variant of gray Hausdorff distance metric was employed to compare images of lesional skin segments of a patient, taken at different time points during a therapeutic course. Prior to image comparison, an effective preprocessing scheme was adapted to constrain wide pose and light variations. The proposed similarity algorithm was tested on raw clinical image data sets of patients diagnosed with toxic epidermal necrolysis, a life-threatening condition with rapid evolution. Fine tuning of algorithm’s parameters was optimized using Precision-Recall curves.

Results: Proposed image comparison method resulted in a high-degree of image similarity (about 96%) between pictures taken at second and fifth day of hospitalization. Current similarity results substantiate a significant agreement between the computer-treatment assessment, by means of image comparison, and the corresponding clinical experts’ review of skin condition.

Conclusion: Objective evidence of ‘no further change’ situation may provide (a) intuitive clinical decision support to dermatologists in assessing aggressive skin conditions, where the timely evaluation of treatment response is of vital importance and (b) a versatile end-point measure for corresponding therapeutic clinical trials.

Key words: Image similarity – Hausdorff Distance – Toxic Epidermal Necrolysis – Treatment evaluation

© 2012 John Wiley & Sons A/S
Accepted for publication 9 August 2012

In medicine, the timely assessment of patient’s response to an ongoing therapeutic intervention is of vital importance for the clinical evaluation of a therapeutic scheme. Particularly in dermatology, the stabilization of the overall skin condition can be in many cases the earliest qualitative measure of the effectiveness of the therapeutic intervention. Subjective image comparisons, that offer empirical ‘qualitative’ judgments of degrees of image similarities, are traditionally employed by the involved physicians.

For medical uses, a quantitative approach to image similarity has been primarily used in image registration problems as a criterion of best matching (1). Currently, with large image databases becoming a reality in many medical domains, image matching is an active research area in content-based image retrieval systems (2–5).

Especially, within the framework of image similarity analysis applications to skin diseases, to the best of our knowledge, most previous work focused on mole evaluation: Several studies have proposed a computerized comparison of serial skin images aiming to assist nevi diagnosis and melanoma identification (6–9). In practice, these previous studies applied matching or ‘registering’ single lesions or selected skin areas in a pair of metachronous images mainly looking for clinically significant changes, i.e. evidence for emerging new lesions or certain alterations in already existing ones.
To date, for computer-vision applications a large number of image similarity metrics, including Hausdorff distance, has been evaluated (10). In the past, Hausdorff distance has been employed with success in many biometric applications like face recognition and video surveillance systems (11–13). Although Hausdorff distance involves very accurate results and low-error rates, a literature review revealed that overall robustness can be improved if the method is accordingly adapted to be used in specific applications (14–16).

In this study, we demonstrate how to adjust gray Hausdorff metrics to the analysis of pathologic skin conditions. Hereby we propose an efficient algorithm capable of reliably identifying degrees of similarity of lesional skin images of a patient, taken in the course of a therapeutic intervention. The main goal was to quantify, by means of an image similarity metric, the degree of stabilization of an expanding skin disease, i.e. to identify the situation of ‘no further change’ of the skin condition of the patient, providing thus the physician with an early, objective measure of the efficacy of the used therapy.

Methods

Skin condition similarity for treatment assessment in toxic epidermal necrolysis
Clinical images of patients diagnosed with toxic epidermal necrolysis (TEN) that underwent an innovative combination therapy are presently compared (17).

TEN is a rare, potentially life-threatening dermatologic disorder, usually resulting from an adverse reaction to medications. It is characterized by widespread erythema, extended epidermal cell necrosis, and detachment of the epidermis and mucous membranes, resulting in massive skin dysfunction and possibly sepsis and death of affected individuals. Moreover, mucous membrane involvement can result in gastrointestinal hemorrhage, respiratory failure, ocular abnormalities, and genitourinary complications (18, 19). The estimated mortality rate associated with TEN varies widely in different reports, from 10% to 70%. The outcome depends in part on the quality of care and the rapidity with which treatment is initiated.

In clinical dermatology, photographic documentation has become an essential element for assessing the impact of therapeutic regimens on a variety of dermatologic conditions. An example is given in Fig. 1, where digital photos document the outcome of a therapeutic medication intervention (for details concerning this specific therapeutic intervention see: (17)).

According to clinical ‘expert-eye’ evaluation, the applied therapeutic trial efficiently arrested the disease progression already from the second to third hospitalization days on. In addition to this subjective expert’s review, it would be useful to be able to automatically analyze such images and offer an objective quantitative estimation of the therapeutic response.

To imitate the expert’s similarity criterion and accurately predict treatment response, an image comparison algorithm has been employed based on gray Hausdorff algorithm has been employed based on gray Hausdorff Distance (12).

A brief overview of Hausdorff Distance
The classical Hausdorff Distance between two point sets \( A = \{a_1, a_2, \ldots, a_m\} \) and \( B = \{b_1, b_2, \ldots, b_n\} \) of finite sizes \( m \) and \( n \), respectively, is given by:

\[
H(A, B) = \max(h(A, B), h(B, A))
\]

(1)

where \( h(A, B) \) and \( h(B, A) \) represent the directed distances between the two sets \( A \) and \( B \). \( h(A, B) \) is defined as:

\[
h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|
\]

(2)

Fig. 1. Efficient treatment has stopped TEN progression; (a) Patient’s baseline presentation (2nd day of treatment) and (b) subsequent treatment outcome (5th day of treatment).
where, $|| \cdot ||$ is some underlying norm on the points A and B. Usually an Euclidian norm is used.

To estimate the directed distance $h(A, B)$, each point of B is ranked based on its distance to the nearest point of A, and the largest ranked point determines the distance. The Hausdorff Distance $H(A, B)$ is the maximum of $h(A, B)$ and $h(B, A)$, estimating thus the degree of mismatch between the two sets by measuring the distance of the point A that is farthest from any point of B and vice versa.

In the case of images, point sets are pixel sets containing the coordinates of feature pixels, for example edges.

The conventional Hausdorff distance, is sensitive to noise. In real applications of image matching, the mismatch will be large when part of an object in an image is missing or when there are outliers. The partial Hausdorff distance (11), a more tolerant to noise metric, is used by taking the K-th ranked maximum value instead the overall maximum in the directed Hausdorff distance:

$$H_p(A, B) = \max(h_p(A, B), h_p(B, A))$$ (3)

Where, $h_p(A, B)$ is the partial directed Hausdorff distance given by:

$$h_p(A, B) = Kth \min_{a \in A, b \in B} \|a - b\|$$ (4)

The Hausdorff metric has interesting features which differentiates it from many other comparison methods; there is no explicit pairing between the point sets A and B and the metric is quite tolerant of small position errors.

Partial Gray Hausdorff Distance (GHD) has been originally introduced by Vivek and Sudha for comparing face images (12). GHD compares image content by comparing distributions of gray values. Assuming we have two images to compare, that is a model image $M$ and a test image $T$ each of them having $n$ different pixel sets. If eight bits are used for representing gray values there will be $n = 256$ different pixel sets in each image. Let $M_0, \ldots, M_n$ and $T_0, \ldots, T_n$ be the pixel sets obtained from images $M$ and $T$, respectively. The $M_i$ and $T_i$ consist of the coordinates of pixels with the gray value $i$. The GHD $H_g(M, T)$ is defined as the maximum of the Hausdorff distances between the corresponding pixel sets of images

$$H_g(M, T) = \max_{i=0..n-1} H(T_i, M_i)$$ (5)

The partial GHD is given by:

$$H_{pg}(M, T) = \max(h_{pg}(M, T), h_{pg}(T, M))$$ (6)

where,

$$h_{pg}(M, T) = Kth \min_{m \in M_i} d(m, T_i)$$ (7)

and

$$d(m, T_i) = \begin{cases} \min_{t \in T_i} ||m - t||, & \text{if } T_i \text{ is non-empty} \\ L & \text{otherwise} \end{cases}$$ (8)

$d(m, T_i)$ gives the distance between $m \in M_i$ and its nearest point in $T_i$. The norm $|| \cdot ||$ is the Euclidean norm. If $T_i$ is empty, the distance is $L$, where,

$$L = \sqrt{r^2 + c^2}$$ (9)

and $r, c$ number image rows and column, respectively.

Instead of giving the value of $K$ directly in Eq. (7), it can be specified as a fraction $f$ of the total number of pixels in the image sets $M$ or $T$. That is, $K = f |M|$ and $K = f |T|$ for $h_{pg}(M, T)$, and $h_{pg}(T, M)$, respectively. $|M|$ denotes the number of pixels in $M$.

Application of GHD for skin condition similarity to monitor treatment effectiveness in patients with TEN

To give quantitative evidence of the state of disease stability, digital images from patients with TEN, taken at the second and fifth days – i.e. end of the interventional phase of the supplied therapy – , were processed and compared using image content similarity criterion based on GHD. Digital body surface pictures of patients were uniformly taken at bed-side from ‘clinically’ appropriate object distance and under conditions of day light illumination using a commercial Olympus X-930 camera (4.65–18.6 mm, 1 : 2.6–5.8, 12 Mpixel) operated on ‘auto’ mode with minimal zoom.

Image preprocessing

Images were converted to gray scale, and an automatic image registration was used to bring
the images to the same coordinate system (20). Spatial transformation inferred from control points manually picked from patient’s body landmarks (nevi, areola) (Fig. 2 and Fig. 3). Following image alignment, images were re-quantized retaining only the most significant bits for perceptual appearance. Using all 256 gray levels increases the computational time of the GHD algorithm without any improvement in its performance accuracy (12). Retaining only the most significant bits of gray value it is possible to tolerate slight changes in lighting conditions. However, in our specific application the image correction for non-uniform illumination using a black top hat morphological filter (20) in conjunction with histogram adjustment operation (20) has significantly improved the algorithm’s performance.

**Computation of GHD**

In this study, to boost the GHD algorithm performance in comparing image content, we employ a variant of partial GHD; instead of estimating the K-th ranked maximum value over all gray values (Eq.7), the partial directed GHD is estimated in two steps:

First, for each gray value $i$ we compute:

$$ h_{pg}^i(M, T) = K_{th} \max_{m \in M_i} d(m, T_i) $$

and then we calculate the variant of partial directed GHD as follows:

$$ h_{pgvar}(M, T) = \frac{N_{th}}{n} \sum_{i=0}^{N-1} h_{pg}^i(M, T) $$

The variant of partial GHD between two images $H_{pgvar}(M, T)$ is given by

$$ H_{pgvar}(M, T) = \max(h_{pgvar}(M, T), h_{pgvar}(T, M)) $$

The higher the partial GHD, the more dissimilar are the images $M$ and $T$. To use a similarity measure in the scale of the range of $[0–1]$, the similarity index is calculated as follows:

$$ \text{similarity} = 1 - \frac{H_{pgvar}(M, T)}{L} $$

The software for image preprocessing and image comparison was developed using MATLAB (R2010a, The MathWorks Inc., Natick, MA, USA).

**Performance evaluation**

The variant GHD similarity algorithm was evaluated in image data acquired by clinical dermatologists. More specifically, the testing data set included 60 image samples corresponding to 30 pairs of lesional skin areas (10 image pairs per case) acquired from three different cases (patients) diagnosed with TEN. Image samples were of size $100 \times 100$ pixels taken at 25% block overlapping from registered patient’s skin images. Each pair of such images corresponded to the appearance of the same body surface region at the second and fifth day of

---

**Fig. 2.** Image acquired in the second day of treatment (a) is considered the Model or reference to which the image acquired in the fifth day of treatment (b)-called Test image- is compared. The object of image registration is to bring the Test image into alignment with the Model image by applying a spatial transformation. Marked points in the pair of images identify the same features or body landmarks. A spatial mapping is inferred from the positions of these control points.

**Fig. 3.** After spatial registration both Model (a) and Test image (b) are in the same coordinate system enabling the comparison of skin condition during treatment course.
treatment. The earliest evidence of a successful treatment outcome was empirically diagnosed by the dermatologists as the situation of ‘no further change’ in the appearance of the skin condition of a patient beyond the second day of treatment.

The power of the variant GHD similarity metric, in identifying true lesion pairs was investigated. For this purpose, two testing lesion sets were considered; the set of true lesion pairs consisting of the 30 sample pairs of ‘No change’ skin condition, and a set of false lesion pairs consisting of 1200 sample pairs of lesion images from different patients.

The performance of the similarity algorithm was evaluated by means of Precision-Recall (PR) curves (21). PR curves have been used in information retrieval as an alternative to Receiver Operator Characteristic (ROC) curves, for recognition tasks with considerable skewed datasets.

Assuming that similarity indexes estimated by variant GHD algorithm can discriminate true from false lesion pairs, then the decision made by the algorithm can be presented in a structure of table known as a confusion matrix (Table 1). True positives (TP) are sample pairs correctly identified as lesion pairs. False positives (FP) refer to actual false pairs incorrectly recognized as true lesion pairs. True negatives (TN) are false pairs correctly identified as false pairs. Lastly, false negatives (FN) refer to true pairs incorrectly recognized as false pairs. In Precision-Recall space, one plots recall on the x-axis and precision on the y-axis. In this study, recall is the true positive rate that measures the fraction of true pairs correctly retrieved by the similarity algorithm. Precision measures that fraction of pairs identified as true pairs that are actual true pairs. Given a confusion matrix (Table 1), precision and recall are estimated as follows:

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (14)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (15)
\]

Correspondingly to ROC curves, the area under the curve (AUC) is also used in PR-curves (AUC-PR) as a measure to define how an algorithm performs over the Precision-Recall space (21). The higher the AUC-PR, the more optimal is the algorithm performance in terms of precision and recall.

The performance of variant GHD algorithm is dependent on three parameters. The parameter \(f_1\) (Eq. 10), which is the fraction of the total number of pixels with gray value \(i\) in the image sets \(M\) or \(T\), the parameter \(f_2\) (Eq. 11) which is the fraction of the total number of \(i\)th directed gray Hausdorff distances \(h_{pg}^{i}(M, T)\) or \(h_{pg}^{i}(T, M)\) and the number \(n\) of pixel value sets. The variant GHD algorithm was evaluated for fraction values \(f_1\) and \(f_2\) ranging from 0.5 to 0.95 with step value 0.05 and for different pixel value sets \(n = (8, 16, 32, 64, 128)\).

**Results**

The selection of parameters \(f_1\), \(f_2\), and \(n\) was optimized on algorithm performance measured by the AUC-PR. The higher AUC-PR value was achieved for \(f_1 = 0.6\), \(f_2 = 0.75\), and \(n = 32\). Table 2 summarizes the algorithm’s performance for different parameter values \(f_1\), \(f_2\) and \(n = 32\).

Figure 4, depicts the variant GHD performance in PR space for optimal parameters' setting. Finally, from the optimal PR-curve (Fig. 4), the best operating point was chosen so that the algorithm could give the best tradeoff between the costs of failing to detect true pairs against the costs of raising false positives. More specifically, having chosen as operating point the similarity value of 0.9553 the variant GHD algorithm recognized correctly the ‘No change’ condition in 25 of 30 true lesion pairs giving a recall of the order of 83.3% at precision level of 92.6% (Table 3). Three FN originated from case 1, one FN from case 2, and one FN from case 3. On average the overall similarity score for the three cases under examination was 0.96.

The estimation of the directed partial GHD for each individual gray value of the tested images has significantly enhanced the efficiency of the algorithm to capture the similarity of

---

**TABLE 1. Confusion Matrix**

<table>
<thead>
<tr>
<th>Actual Positive (True lesion pairs of ‘No change’ Condition)</th>
<th>Actual negative (False pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted True pairs</td>
<td>Predicted False pairs</td>
</tr>
<tr>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

---
Comparison results employing the original partial GHD algorithm \((12)\) for varying parameters \(f\) and \(n\), are given in Table 4. Optimal performance of original GHD algorithm was achieved for \(n = 64\) and \(f = 0.9\) (Table 4). Chosen as operating point the similarity value of 0.9588 the original GHD algorithm yielded a recall of the order of 80\% at precision level of 80\% (Fig. 5).

It is noticeable, that since the GHD algorithm operates directly on image gray values, its performance might be highly affected from the image preprocessing procedure. In the present application, the images have been acquired in a clinical routine setting which means that they might have been taken in a suboptimal illumination condition or might be affected by wide illumination variations. Probably, for such reasons, the GHD algorithm yielded a remarkably low performance when applied directly to the original images (Table 5, Fig. 6). Image enhancement using black top hat morphological filter in conjunction with histogram adjustment operation has been proved an essential preprocessing step for the application of GHD algorithm in this study.

### Discussion

In this study a methodology has been proposed to measure pair wise the macromorphological similarity of meta-chronically acquired clinical images of the same body surface region. Our effort was to quantify the visual measure of the degree of stabilization (‘no further change’, i.e. no progression) of the clinical condition. This would be of great value in clinical dermatology, enabling experts to quantify the impact of therapeutic regimens on a variety of dermatological diseases, including some potentially life-threatening ones.

Development of new lesions and/or enlargement of existing ones are in most cases the presentation forms of a still progressing skin disease. Accordingly, the clinical judgment of effectiveness of a therapeutic intervention is
based on the observation of tissue alterations signifying return to the prior physiological situation. Especially in the case of skin conditions with relatively slow tissue damage restitution rate, the observation of the ‘freezing’ of the morphological skin alterations is the earliest criterion evidencing treatment success. Intuitively under such circumstances, the earliest clinical indication of achieving the goal of treatment is the manifestation of no further alteration of patient’s clinical state, which is equivalent to similarity of clinical images taken at appropriate time points.

In general, many pathological skin conditions that cause significant structural skin damage that heal relatively slowly, fulfill the conditions to be analyzed this way. Besides TEN, such skin diseases include many forms of vasculitis, blistering skin diseases, and pigmentation alterations but also different forms of viral or drug-induced eruptions. Notably, the same principles of evaluating image similarity are widely applied in many different medical fields, for example in radiology, to appreciate the nature of pathological findings on the basis of their morphological evolution.

In the present study, quantification of disease arrest involved the individual comparison of lesion pair samples taken as image pair samples at selected time points during hospitalization. The proposed method was evaluated on raw clinical image data sets of patients diagnosed with TEN to support early treatment assessment.

In everyday clinical practice it is rarely feasible to take standardized pictures as this can be time-consuming and patients cannot always co-operate (elderly), or like TEN their condition does not allow for manipulations (pain, risk of infection). To constrain wide pose and light variations of taken photographs, prior to images comparison, an effective preprocessing scheme was adapted.

### TABLE 4. Comparative AUC-PR results for original partial GHD algorithm for different values of parameters \( n \) and \( f \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( f )</th>
<th>0.5</th>
<th>0.55</th>
<th>0.6</th>
<th>0.65</th>
<th>0.7</th>
<th>0.75</th>
<th>0.8</th>
<th>0.85</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.459263</td>
<td>0.461775</td>
<td>0.428096</td>
<td>0.487803</td>
<td>0.466463</td>
<td>0.522825</td>
<td>0.577699</td>
<td>0.662759</td>
<td>0.689187</td>
<td>0.698123</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.495371</td>
<td>0.485316</td>
<td>0.587376</td>
<td>0.706222</td>
<td>0.750902</td>
<td>0.764457</td>
<td>0.813375</td>
<td>0.798609</td>
<td>0.817958</td>
<td>0.745495</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.514095</td>
<td>0.542283</td>
<td>0.624705</td>
<td>0.688219</td>
<td>0.767093</td>
<td>0.793156</td>
<td>0.818225</td>
<td>0.825091</td>
<td>0.813633</td>
<td>0.716649</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.582658</td>
<td>0.584827</td>
<td>0.629252</td>
<td>0.615644</td>
<td>0.782307</td>
<td>0.81081</td>
<td>0.834799</td>
<td>0.815447</td>
<td>0.817863</td>
<td>0.709904</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.571132</td>
<td>0.664127</td>
<td>0.688421</td>
<td>0.798418</td>
<td>0.801927</td>
<td>0.813749</td>
<td>0.826535</td>
<td>0.83433</td>
<td>0.83589</td>
<td>0.693296</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>0.613462</td>
<td>0.67039</td>
<td>0.776308</td>
<td>0.788197</td>
<td>0.818076</td>
<td>0.802982</td>
<td>0.805571</td>
<td>0.78181</td>
<td>0.748032</td>
<td>0.522998</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 5. Original GHD performance in PR space for optimal parameters setting (\( n = 64, f = 0.9 \)). Chosen as operating point the similarity value of 0.9588 the original GHD algorithm yielded a recall of the order of 80% at precision level of 80%.

### TABLE 5. Comparative AUC-PR results for partial GHD algorithm for different values of parameters \( n \) and \( f \) applied to original image samples

<table>
<thead>
<tr>
<th>( n )</th>
<th>( f )</th>
<th>0.5</th>
<th>0.55</th>
<th>0.6</th>
<th>0.65</th>
<th>0.7</th>
<th>0.75</th>
<th>0.8</th>
<th>0.85</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.489417</td>
<td>0.473852</td>
<td>0.502437</td>
<td>0.505143</td>
<td>0.508525</td>
<td>0.480488</td>
<td>0.441857</td>
<td>0.407307</td>
<td>0.430623</td>
<td>0.239747</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.517682</td>
<td>0.521782</td>
<td>0.507669</td>
<td>0.5406</td>
<td>0.505493</td>
<td>0.517157</td>
<td>0.391594</td>
<td>0.348238</td>
<td>0.332575</td>
<td>0.114888</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.554087</td>
<td>0.563781</td>
<td>0.535572</td>
<td>0.521646</td>
<td>0.497172</td>
<td>0.476222</td>
<td>0.369196</td>
<td>0.270001</td>
<td>0.118617</td>
<td>0.123657</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.559626</td>
<td>0.508447</td>
<td>0.490494</td>
<td>0.491276</td>
<td>0.459294</td>
<td>0.419193</td>
<td>0.283456</td>
<td>0.217779</td>
<td>0.115765</td>
<td>0.089362</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.383675</td>
<td>0.370844</td>
<td>0.287253</td>
<td>0.207792</td>
<td>0.11591</td>
<td>0.071333</td>
<td>0.045352</td>
<td>0.016828</td>
<td>0.016592</td>
<td>0.015802</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>0.056029</td>
<td>0.017447</td>
<td>0.017155</td>
<td>0.016293</td>
<td>0.012195</td>
<td>0.012195</td>
<td>0.012195</td>
<td>0.012195</td>
<td>0.051295</td>
<td>0.512195</td>
<td></td>
</tr>
</tbody>
</table>
The variant GHD similarity algorithm was effectively tuned and tested for its ability to recognize the disease stability. More specifically, employing the variant GHD metric, and for similarity values greater than 0.9553, the algorithm was able to recognize the spatial stability of the eruption with an accuracy in the order of 83.3% and with a confidence of 92.6%, as implied by the fact that a similarity index \( C_2 \) \( \geq 0.9553 \) is expected to originate from a corresponding image content constellation. Finally, overall case evaluation was summarized by the average similarity value estimated from image samples. For all three TEN cases tested in this study the variant GHD algorithm computed an overall similarity of 0.96 of clinical images in due of the therapy, which corresponds to the clinical experts’ impression of an efficacious treatment modality.

**Conclusion**

In this study, an effective method was implemented to process and compare images in the setting of everyday clinical dermatology. Our aim was to quantify the earliest treatment effects of therapeutic regimens by measuring the degree of disease stabilization using image similarity analysis. Our approach involved the comparison of images of selected lesional skin segments of the same patient, taken at different time points during a therapeutic course. The image similarity criterion was based on a variant of gray Hausdorff distance metric. The proposed similarity algorithm exhibited significant agreement with expert’s ‘clinical’ evaluation of treatment efficacy, suggesting the reliability of the employed method in comparing images of lesional skin. Since the image similarity criterion is based on image gray values and not on specific lesion features, our treatment evaluation approach could be used in many pathological skin conditions that cause significant structural skin damage and heal relatively slowly. In conclusion, objective recognition of the status of ‘no further spatial change’ in the course of the treatment of a presumably expanding pathological skin condition may provide: (a) intuitive clinical decision support to dermatologists in assessing the response of skin diseases, especially life-threatening aggressive ones, to the provided therapeutic modality and (b) a versatile end-point measure for corresponding future therapeutic clinical trials.

**References**


Address:
Panagiota Spyridonos
Department of Medical Physics
University of Ioannina Medical School
University Campus
45110, Ioannina
Greece
Tel: +302651007782
Fax: +302651007854
e-mail: pspyrid@cc.uoi.gr