DIMACS Technical Report 2004-53
November 2004

Optimal Mapping of Deep Gray Scale Images to a Coarser Scale of Gray

by

Solomon Borodkin
CACI International, Inc.
4831 Walden Lane, Lanham, MD 20706
sborodkin@caci.com

Aleksey Borodkin
BAE Systems,
2 Massachusetts Ave, Washington, DC 20212
Aleksey.Borodkin@baesystems.com

Ilya Muchnik
DIMACS
Rutgers University
P.O. Box 8018, Piscataway, NJ 08855
muchnik@dimacs.rutgers.edu

Abstract

The problem of optimal mapping of multi-level gray scale images to a coarser gray scale is formulated and explored. A distance between images having different intensity ranges is introduced. Minimization of such a distance can be viewed as least squares approximation of a source high range image by the best target image with a given number of levels of gray. Following S.Lloyd [1], we proved that the latter problem is equivalent to optimal partitioning of the source image intensity range into a given number of intervals, provided that the sum of intra-interval variations reaches minimum. An efficient algorithm for optimal partitioning based on dynamic programming is used, which is the same in complexity as the ones known from literature, but better in terms of required memory. The proposed approach is applied to visualization of deep gray scale medical images. Advantages of the method over linear mapping and histogram equalization are demonstrated on the sample images. The other application fields may include image optimization for printing/faxing/copying.

1 Introduction

Medical imaging equipment (CT- and MRI-scanners, X-ray, etc.) produces images of intensity range 9 - 13 bits per pixel, while conventional image presentation devices and media provide a lower range. Personal computers usually allow for only 8 bits, or 256 levels of gray. In this paper, we introduce a problem statement for optimal mapping of deep gray scale images to a coarser gray scale – optimal gray scale requantization (*OGSR*). It is based on the best approximation of a fine source image by its coarser version from a class of images with the limited number of shades of gray.

We prove that such an approximation is equivalent to optimal partitioning of the source intensity range into a given number of intervals, provided that the average intra-interval scattering reaches minimum. Thus, for our specific problem, we re-created a result derived earlier for a different data model in the signal quantization theory [1].

Sub-optimal heuristic partitioning procedures were proposed in a number of works (e.g., [1]-[6]). While these procedures are usually simple, fast, and acceptable in most of cases, they don't provide global optimum, i.e. don't guarantee a reasonable output for any input image (see [7]). On the other hand, a globally optimal and efficient algorithm of scalar segmentation, based on dynamic programming, was proposed in [8]-[10]. In our paper, we present an algorithm equivalent to [10] in terms of potential complexity and performance, yet better in terms of required memory.

According to our method, intensity of each pixel in the target image is equal to the ordinal number of the interval containing intensity value of the corresponding pixel in the source image. As a result, the variably spaced source intensity intervals are mapped to equally spaced destination intensity values. It is not clear *a priori* whether such a non-linear transformation would yield still recognizable pictures; sample images in Section V can be a cogent argument *pro*. The optimally mapped images are significantly more realistic than those obtained by the histogram equalization (*HE*) technique.

Attention is also drawn to practical aspects of medical image presentation. In order to take into account contrast distortions due to non-linearity of intensity ranking, we propose a "double-view" displaying: linearly mapped (*LM*) image together with the one obtained by the *OGSR* method; that makes it possible to view maximum of image details while referring the true shades of gray.

The method also can be used for pre-processing images aimed at high quality printing or faxing.

2 Problem statement

Consider source image A:$\{\hat{a}_1,...,\hat{a}_M\}$ as an ordered set of M pixels; each pixel $\hat{a}_i : \{x_i, y_i; a_i\}$ is defined by its coordinates and intensity value a_i. The order of pixels is assumed fixed, so image A can be identified by its intensity vector $\{a_1,...,a_M\}$ in the M-dimensional space. Components a_i are assumed, for simplicity, real non-negative numbers, which define a *source intensity scale*.

Let B:$\{b_1,...,b_M\}$ denote another image – a vector with components evaluated in the *target scale*, which is a set of integers in $\overline{0, n-1}$.

Given source image A, in order to find its best presentation in the target scale, we need to define distance between the two images with the intensity values belonging to different scales.

Consider a set of n monotonically increasing parameters β^k - values in the source scale, such that:

$$0 \le \beta^0 \le \beta^1 \le ... \le \beta^{n-1} \le \max_{\{1 \le i \le M\}}(a_i) = a^{(\max)} \tag{1}$$

Let D define distance between A and B as follows:

$$D(A, B) = \min_{\{\beta^0,...,\beta^{n-1}\}} \sum_{i=1}^{M} (a_i - \beta^{b_i})^2 \tag{2}$$

The major problem of this work is finding the best approximation of A in the target scale:

$$B_A = \arg\min_{\{B\}} D(A, B) \tag{3}$$

Let Ψ^r denote a class of all images that have equal or less than *r* distinct levels of gray. Obviously, *A* belongs to a certain Ψ^N, where $N \le M$, while $B \in \Psi^n$; it is assumed $n < N$.

Let also $\overline{B} : \{ \overline{b}_1, ..., \overline{b}_M \}$ be an image in the source scale, defined by image *B* and parameter set $\{ \beta^k \}$, so that $\overline{b}_i = \beta^{b_i}$. Note that although the absolute values of gray are different in $\overline{B}$ and *B*, both belong to Ψ^n, and their *equi-intensity regions* (EIRs) are congruent.

From (2) and (3), obtain a simpler form for *D*(*A*, B_A) criterion value:

$$\min_{\{\overline{B} \in \Psi^n\}} \sum_{i=1}^{M} (a_i - \overline{b}_i)^2 \qquad (4)$$

Thus, optimal image mapping to a lower gray scale can be split into the two steps $A \to \overline{B} \to B$:

- For a given source image *A* find the best (least squares) approximation $\overline{B} \in \Psi^n$ according to (4);
- Sort the obtained distinct intensity values of $\overline{B}$ in ascending order, starting from β^0, and define B_A (optimal *B* for a given *A*) by setting $b_i = k$, if $\overline{b}_i = \beta^k$.

Intensity values provide a two-fold differentiation of the parts of the whole image: first, they define EIRs as disjoint regions in the picture (*spatial details*); second, they define a specific scalar attribute to each EIR – absolute value of gray (*contrast details*). During $A \to \overline{B}$ step, we preserve as much spatial and contrast details as possible. During $\overline{B} \to B$ step, we preserve all the spatial details but introduce additional contrast distortion due to mapping the variably spaced values of $\overline{B}$ to equally spaced values of *B* (only ordering of gray shades is conserved).

In the following two sections, we explore the $A \to \overline{B}$ step.

3 Image approximation and partitioning of intensity range

Expression (4) defines a least squares image approximation problem, close to quantization of random signal. If re-formulated in terms of our paper, it was shown that under certain conditions, problem (4) could be reduced to the following one:

find (n–1) values { $a^1,\ldots,a^{n-1}$ *} and n values {* $\beta^0,\ldots,\beta^{n-1}$ *} - totally (2n–1) values in the source scale,*

$$0 = a^0 \le \beta^0 < a^1 \le \beta^1 < \ldots < a^{n-1} \le \beta^{n-1} < a^n = \max(a_1,\ldots,a_M)+1 \tag{5}$$

provided:

$$\min_{\{a^r,\beta^r;\, r\in\overline{0,n-1}\}} \sum_{k=0}^{n-1}\left[\sum_{a^k\le a_i<a^{k+1}} (a_i-\beta^k)^2\right] \tag{6}$$

In other words, split the whole intensity range [a^0, a^n), with a^0 inclusive, a^n non-inclusive, by (*n*-1) points { a^k; $k \in \overline{1,n-1}$ } into *n* semi-open intervals, and find certain value β^k within each interval *k*, - to provide minimum of intra-interval intensity variations.

In [1], the “(4) $\Rightarrow$ (6)” reduction was obtained under the different conditions: continuous space, probabilistic model with additive noise. Even more important is that a heuristic algorithm used for quantization did not require accurate treatment of what was viewed as situations of probabilistic “measure zero” (where pixel intensities fell into endpoints of the intervals). So some of the situations, common to our problem, were ignored in [1].

In fact, reduction (4) $\Rightarrow$ (6) turns out to be true, as well, even if the mentioned situations have a non-zero probability, like it is the case in the image approximation; - moreover, *equivalence*: (4) $\Leftrightarrow$ (6).

Prior to formulating the equivalence theorem, we introduce another definition. Given images *A* and $\overline{B}$, define a set of disjunctive regions { $A^0,\ldots,A^{n-1}$ } of *A*, where A^k is a set of all pixels $\hat{a}_i$ such that $\hat{a}_i \in A^k$ if and only if $\overline{b_i} = \beta^k$:

$$A^k : \{\hat{a}_i \mid \overline{b_i} = \beta^k; i \in \overline{1,M}\}, \quad 0 \le k < n \tag{7}$$

The following Theorem (proved in Appendix) can be viewed as a more accurate re-creation of the S.Lloyd’s result [1], if applied to image approximation.

Theorem

If image $\overline{B} \in \Psi^n$ *is the best approximation of source image* $A \in \Psi^N$ *in terms of least squares* (4)*, then:*

- *Regions* A^k *defined by* (7) *are strictly ordered by intensities, i.e.*

$a_i < a_j$ *for any pixels* $\hat{a}_i \in A^p$, $\hat{a}_j \in A^q$ *(0 ≤ p<q < n).*

- *For any* k *(0 ≤ k < n), value* β^k *is an average of intensity values of pixels in region* A^k*, and there exists exactly n different intensity values in set* $\{\beta^k; k=\overline{0,n-1}\}$*:*

$$0 \le \beta^0 < \beta^1 < \ldots < \beta^{n-1} \le \max_{\{1\le i\le M\}}(a_i) \tag{8}$$

Theorem states that there is a one-to-one correspondence between regions A^k and intervals $[a^k, a^{k+1})$ of the source scale - this is why expressions (4) and (6) are equal (which proves "(4) ⇒ (6)"):

$$\min_{\{\overline{B}\in\Psi^n\}} \sum_{i=1}^{M}(a_i - \overline{b}_i)^2 = \min_{\{A^r,\beta^r;r=0,n-1\}} \sum_{k=0}^{n-1} \sum_{\hat{a}_i\in A^k}(a_i - \beta^k)^2 = \min_{\{a^r,\beta^r;r\in 0,n-1\}} \sum_{k=0}^{n-1}\left[\sum_{a^k\le a_i<a^{k+1}}(a_i - \beta^k)^2\right] \tag{9}$$

Based on the Theorem and (9), it is easy to prove the converse assertion, "(6) ⇒ (4)": if intervals $[a^k, a^{k+1})$ and values β^k ($k\in \overline{0,n-1}$) provide minimum (6), then the corresponding image $\overline{B}$ with components $\overline{b}_i = \beta^k$ where k is defined from $a_i \in [a^k, a^{k+1})$, - provides minimum (4).

Switching from (4) to (6) is a significant simplification, because the number of unknown variables is reduced from M in (4) to 2*n*-1 in (6), where usually n<<M.

Note that problem (6) may be interpreted as another least squares approximation problem: *find the best approximation of the source image intensity histogram by the n-step-wise function* [10]. It is different from the original approximation problem (4); however, according to the Theorem, both are equivalent in the following sense: given a solution for one of them, we can easily get a solution for the other one.

4 Optimal partitioning by dynamic programming

Let m_g be multiplicity of intensity g in the source image, i.e. a number of pixels with intensity g. In order to simplify notation, from now, we will assume that source values of gray are integers in $\overline{0, N-1}$.

We also assume that although some of source “intensity slots” may be empty (m_g = 0), the number of distinct levels of gray (with $m_g > 0$) is greater than *n*.

A sequence { $m_0, m_1, ..., m_{N-1}$ } defines the source image histogram; so the sum to be minimized in (6) can be reduced to:

$$\sum_{k=0}^{n-1}\sum_{g=a^k}^{a^{k+1}-1} m_g\left[g-\left(\sum_{r=a^k}^{a^{k+1}-1} m_r r \Big/ \sum_{r=a^k}^{a^{k+1}-1} m_r\right)\right]^2 = \sum_{k=0}^{n-1}\left[\sum_{g=a^k}^{a^{k+1}-1} m_g g^2 - \left(\sum_{g=a^k}^{a^{k+1}-1} m_g g\right)^2 \Big/ \sum_{g=a^k}^{a^{k+1}-1} m_g\right] = \sum_{k=0}^{n-1} f(a^k, a^{k+1}),$$

(10)

where

$$f(g,r) = \sum_{p=g}^{r-1} m_p p^2 - \left(\sum_{p=g}^{r-1} m_p p\right)^2 \Big/ \sum_{p=g}^{r-1} m_p, \quad (0 \le g < r \le N) \tag{11}$$

By using three one-dimensional arrays Q_g, S_g, H_g ($0 \le g \le N$), which can be calculated in advance:

$$Q_g = \sum_{p=0}^{g-1} m_p p^2, \; S_g = \sum_{p=0}^{g-1} m_p p, \; H_g = \sum_{p=0}^{g-1} m_p, \quad (0 < g \le N); \quad Q_0=S_0=H_0=0, \tag{12}$$

we can avoid repetitive summation in (11) when calculating values of $f(g,r)$, or avoid using a large two -dimensional array of (*N*(N-1)/2*) double-precision numbers calculated in advance (like, e.g., in [11]):

$$f(g,r) = (Q_g - Q_r) - (S_g - S_r)^2 / (H_g - H_r), \qquad (0 \le g < r \le N), \tag{13}$$

As a result, criterion (6) gets the following canonical form:

$$D = \min_{\{a^1,a^2,...,a^{n-1}\}} \sum_{k=0}^{n-1} f(a^k, a^{k+1}), \qquad (0 = a^0 < a^1 < a^2 < ... < a^{n-1} < a^n = N) \tag{14}$$

Vector minimization in (14) can be split into a sequence of simpler *scalar* optimization procedures as shown below. Instead of designation D for the criterion value, we will use more general $\tilde{D}^0(a^0)$, - to indicate that the final criterion value is just a special case of a whole family of similar criteria $\tilde{D}^k(a^k)$, which should be calculated in order to obtain the final value of $D = \tilde{D}^0(a^0)$:

$$D = \tilde{D}^0(a^0) = \min_{\{a^1,a^2,\ldots,a^{n-1}\}} \sum_{k=0}^{n-1} f(a^k,a^{k+1}) = \min_{\{a^1\}} \left[f(a^0,a^1) + \min_{\{a^2,a^3,\ldots,a^{n-1}\}} \sum_{k=1}^{n-1} f(a^k,a^{k+1}) \right] =$$

$$= \min_{\{a^1\}} \left[f(a^0,a^1) + \tilde{D}^1(a^1) \right] = \min_{\{a^1\}} \left[f(a^0,a^1) + \min_{\{a^2\}} \left[f(a^1,a^2) + \tilde{D}^2(a^2) \right] \right] = \ldots \quad (15)$$

To resolve (15), we can use the following system of recurrent functional equations of dynamic programming (starting from the last one):

$$\tilde{D}^{n-1}(g) = f(g,N), \qquad (0 \le g < N); \quad (16)$$

$$\tilde{D}^k(g) = f(g,a_g^{k+1}) + \tilde{D}^{k+1}(a_g^{k+1}), \text{ where } a_g^{k+1} = \arg\min_{\{r\}} \left[f(g,r) + \tilde{D}^{k+1}(r) \right] \quad (17)$$

By tabulating (17) for *k*=*n-2,…,1*; *g*=0,…,*N*-*k*-*2*, and populating array a_g^k, we get $D = \tilde{D}^0(0)$. Then, starting from *g=0*, *k=0*, and using array a_g^k, - all the end points of optimal intervals can be restored.

The basic time-consuming part of this algorithm is tabulation in (17), requiring CnN^2 operations, - the same as in [8]-[10], where *C* is a constant, which depends on implementation. Mapping images from *N*= 2048 to *n*=256 levels of gray took several seconds on the 2 *GHz* PC.

5 Examples

The *OGSR* method preserves maximum of spatial image details at the expense of contrast distortion. In medical applications, the absolute levels (or, at least, their proportions) are usually important and cannot be changed or ignored. To reconcile both requirements (true contrast and high image detail resolution), we propose the following *double-view* presentation (i.e. show pairs of pictures):

(a) the view obtained by simple linear mapping (*LM*),

(b) the view obtained by *OGSR* as described in this paper.

The first one is a referential view: it provides true shades of gray but lacks spatial details; the second one makes it visible spatial details by using artificial shading.

Below, we present two deep gray scale medical images downloaded from the Web site of *Mallinckrodt Institute of Radiology of Washington University, St. Louis, Missouri*. We do not discuss medical interpretation of images, — that might be a topic of the special research. Instead, we demonstrate general capabilities of the method, drawing attention to additional image details that become visible.

The results are presented in the target scale 0..255. In addition to the recommended "Double-view" (*a*)-(*b*), we also for comparison, demonstrate one more view *(c)*, obtained by optimal *HE* [12].

Fig. 1 shows sample medical image ***MR-ANGIO***. View (*b*) provides significantly higher quality of image details, than view (*a*). Not only is the vascular net better visible in (*b*), but also more details in the major vessels' walls are resolved. View (*c*) hides image details in the brightest (white) area, which are clearly visible in view (*b*), and even in view (*a*). That is because *HE* maps the sparse brightest intensity values into to a single target intensity value — to provide equally populated intensity intervals in the equalized image. Fig. 2 provides more insight by showing histograms for this image.

In Fig. 2a, the original intensity range of image *MR-ANGIO* is split into 256 optimal intervals (in terms of (6)). The intervals are shown as alternating white and light gray vertical stripes; each interval starts from the first gray value falling in the interval, and spans until the next interval starts. The top half of the diagram presents the original image histogram; the bottom half (with the vertical axis directed downwards) presents the 256 peaks-wise histogram of the corresponding best approximation image $\overline{B}$. The peak-wise nature of the bottom histogram is clearly visible in the right part, where the resolution is sufficient to identify a single peak in each interval.

In Fig. 2b, the top part is the same source image histogram, while the bottom part is a peak-wise histogram corresponding to the *HE* view (each peak is positioned at the interval's average intensity). The alternating vertical white and light gray stripes show the corresponding intensity intervals built by *HE*.

Comparing the bottom histograms in Fig. 2a and Fig. 2b one can see that the *HE* peaks are virtually equal in size, while the *OGSR* peaks closely follow the shape of the source histogram. This is why the *OGSR* image in Fig. 1*b* looks more realistic than the *HE* image in Fig. 1*c*.

Also note that the brightest, poorly populated intensity levels are resolved in Fig 2*a* by about 20 intervals, while in Fig. 2*b*, all the values of the right half of the intensity range fall into only 3 intervals. This is why the brightest image details in the equalized image in Fig. 1*c* are lost.

These observations are typical. Fig. 3 presents another example, ***MR-SHOULDER.*** The source image intensity values fall into range 0..595. (Since image size is 1024×1024, only a part of it is shown in the frame window 256×256; however, the whole range of gray is presented in this visible part). The source intensity range is only twice as wide as the target range; that is why simple linear mapping in view (*a*) provides sufficient display quality. View (*b*) provides further visualization improvement, e.g., some details surrounding the bone head are better visible than in view (a). It is more realistic compared to view (*c*) where details in the brightest area are lost.

6 Conclusion

A method of mapping a deep gray scale image to a lower resolution scale is proposed. The method makes it possible to display images with thousands levels of gray on the regular monitors. It is a fully automatic and image content independent technique, free of heuristic elements, providing the best image presentation in terms of the explicitly formulated and intuitively justified criterion. Computational complexity of the mapping algorithm is $O(nN^2)$, where N is the source image intensity range, and n is the target intensity range.

The presented image examples demonstrate high quality of the obtained views. Despite the essential non-linearity of the method, it causes less “non-linear distortions” than, e.g., *HE*. It provides a robust and virtually realistic view, which appears rather close to the linear mapping while significantly richer in terms of general appearance and detail resolution.

Appendix

Proof of Theorem.

Let A^k stand for a group of pixels (7), corresponding to intensity value β^k in $\overline{B}$.

1). First, show that optimal image $\overline{B}$ has exactly *n* different levels of gray. If minimum in (4) is reached with only *m* levels (*m*<*n*), then at least one of A^k contains pixels with different intensities. By splitting this group, we can further lessen the criterion- in contradiction with (4).

Second, β^k must be an average intensity in A^k. Let $D = \|A - \overline{B}\| = (a_1^k - \beta^k)^2 + \ldots + (a_{p_k}^k - \beta^k)^2 + R$, where R is independent of β^k, hence from $\partial D / \partial \beta^k = 0$, obtain $\overline{\beta}^k = (a_1^k + \ldots + a_{p_k}^k) / p_k$.

2). Now, prove that based on (4), pixel groups A^k define disjoint semi-open intervals of the source intensity range, which have empty intersections with each other.

Consider any two intensity levels of image $\overline{B}$, - for a certainty, β^0 and β^1, ($\beta^0 < \beta^1$).

Without limitation of generality, we can assume that pixels in each of groups A^0, A^1 are ordered by intensity values: $a_1^0 \le a_2^0 \le \ldots \le a_p^0$; $a_1^1 \le a_2^1 \le \ldots \le a_q^1$, (to simplify notation, we use p instead of p_0, and q instead of p_1).

We need to prove that $a_p^0 < a_1^1$. We will show that any alternative would result in contradictions.

2a). The first opposite case $a_p^0 > a_1^1$ can be excluded in a similar way as in [1], as follows.

Let $D = \|A - \overline{B}\| = (a_1^0 - \beta^0)^2 + \ldots + (a_p^0 - \beta^0)^2 + (a_1^1 - \beta^1)^2 + \ldots + (a_q^1 - \beta^1)^2 + S$,

where S stands for all the terms in the sum that are independent of both β^0 and β^1. If we just interchange pixels a_p^0 and a_1^1, which means replacing $\overline{B}$ by $\overline{B}'$ so that $\overline{b}'^0_p = \beta^1$, $\overline{b}'^1_1 = \beta^0$, - the new criterion value is: $D' = \|A - \overline{B}\| = (a_1^0 - \beta^0)^2 + \ldots + (a_{p-1}^0 - \beta^0)^2 + (a_1^1 - \beta^0)^2 +$

$(a_p^0 - \beta^1)^2 + (a_2^1 - \beta^1)^2 + \ldots + (a_q^1 - \beta^1)^2 + S$, and D'-D=$2(\beta^1 - \beta^0)(a_1^1 - a_p^0)$.

The last expression is negative, which means that $\overline{B}'$ is the better approximation than $\overline{B}$, in contradiction with our basic assumption.

2b). Consider the second alternative: $a_p^0 = a_1^1$.

This is the situation, which was assumed in [1] to be of probabilistic measure 0, hence it was ignored. In our problem - in probabilistic terms - it is a regular case, which can't be ignored. The following Lemma is used to prove that still $a_p^0 = a_1^1$ is *not* possible, - however, for quite different reason.

Lemma.

Adding one pixel of intensity x to a group of p pixels A^k (for a certainty, A^0) causes a non-negative change of this group's contribution D^0 to the distance $\|A - \overline{B}\|$ by value Δ:

$$\Delta = (S^0 - px)^2 / [p(p+1)], \text{ where } S^0 = a_1^0 + \ldots + a_p^0.$$

Indeed, let $Q^0 = (a_1^0)^2 + \ldots + (a_p^0)^2$, then $D^0 = (a_1^0 - \beta^0)^2 + \ldots + (a_p^0 - \beta^0)^2 = Q^0 - (S^0)^2 / p$.

After adding a pixel with intensity x, obtain $\overline{D}^0 = Q^0 + x^2 - (S^0 + x)^2 / (p+1)$; by subtracting D^0 from $\overline{D}^0$ get Δ.

The following two corollaries result from the Lemma:

C1). *Adding a pixel to a group never reduces the group's contribution to the distance; the increase is* 0 *only if the added intensity is equal to the mean:* $x = S^0/p$.

C2). *Removing pixel $\hat{x} \in A^0$ from group A^0 (if it is not a single member in A^0) never increases the group's contribution; the decrease is $(S^0 - px)^2 / [p(p-1)]$, and it is* 0 *only if* $x = S^0/p$.

Using these corollaries, we can make sure that by moving one pixel from A^0 to A^1 (or from A^1 to A^0), the presumably minimal distance can be further decreased, which is absurd. Several cases should be considered; for the sake of brevity, we omit trivial cases, e.g. when one of the groups A^0, A^1 does not contain pixels with intensity different from $a_p^0 = a_1^1$, - these cases can be excluded easily.

Consider more typical situation $a_1^0 \leq \ldots < \ldots \leq a_p^0 = a_1^1 \leq \ldots < \ldots \leq a_q^1$ (both groups have members with intensity values different from $a_p^0 = a_1^1$), and try to move $\hat{a}_1^1$ from A^1 to A^0. Increase of distance due to adding $\hat{a}_1^1$ to A^0 is $\Delta D^0 = (S^0 - p\, a_1^1)^2 / [p(p+1)]$; decrease of distance due to removing $\hat{a}_1^1$ from A^1 is $\Delta D^1 = (S^1 - q\, a_1^1)^2 / [q(q-1)]$. If $\Delta D^0 < \Delta D^1$ then we obtained contradiction; so suppose $\Delta D^0 \geq \Delta D^1$. We will show that in this case, moving $\hat{a}_p^0$ from A^0 to A^1 would reduce the distance, which is absurd. Consider the following chain of estimations:

$(S^0 - p\, a_p^0)^2 / [p(p-1)] = (S^0 - p\, a_1^1)^2 / [p(p-1)] > (S^0 - p\, a_1^1)^2 / [p(p+1)] \geq (S^1 - q\, a_1^1)^2 / [q(q-1)] = (S^1 - q\, a_p^0)^2 / [q(q-1)] > (S^1 - q\, a_p^0)^2 / [q(q+1)]$, -

it results in $(S^0 - p\, a_p^0)^2/[p(p-1)] > (S^1 - q\, a_p^0)^2/[q(q+1)]$, which means that moving $\hat{a}_p^0$ from A^0 to A^1 would decrease the value of criterion, - in contradiction with the initial assumption, *Q.E.D.*

LITERATURE

[1]. Lloyd, S.P. "Least Squares Quantization in PCM", *IEEE Trans. On Information Theory*, **IT-28**(2), pp. 129-137, March 1982.

[2]. Weszka, J. and Rosenfeld A. "Histogram modification for threshold selection", *IEEE Trans. Syst. Man Cybernet*. **9**, 38-52, 1979.

[3]. Reddi, S.S., Rudin, S.F., and Keshavan, H.R. "An Optimal Multiple Threshold Scheme for Image Segmentation", *IEEE Trans. Systems, Man and Cybernetics*, **14**(4), pp. 661-665, July 1984.

[4]. Bhattacharya P., Yan Y.-K. "Iterative histogram modification of gray images", *IEEE Trans. Systems, Man and Cybernetics*, **25**(3), pp. 521-523, March 1995.

[5]. Khotanzad A., Bouarfa A. “Image segmentation by a parallel, non-parametric histogram based clustering algorithm.” *Pattern Recognition,* **23**(9), pp. 961-973, 1990.

[6]. Tsai D.-M. and Chen Y.-H. “A fast histogram-clustering approach for multi-level thresholding”, *Pattern Recognition Letters*, **13**, pp. 245-252, April 1992.

[7]. Lee, H., Park R.-H., Comments on “An Optimal Multiple Threshold Scheme for Image Segmentation”, *IEEE Trans. Systems, Man and Cybernetics*, **20**(3), p. 741, May/June 1990.

[8]. Fisher, W.D. “On grouping for maximum homogeneity”, *J.Am.Stat.Assoc*. **53**, 789-798 (1958).

[9]. Borodkin, S.M. “Optimal Grouping of Interrelated Ordered Objects”, *Automation and Remote Control*, Plenum Publishing Corporation, 1980, pp. 269-276.

[10]. Kundu, S. “A solution to histogram-equalization and other related problems by shortest path methods”, *Pattern Recognition*, **31**(3), pp. 231-234, March 1998.

[11]. Borodkin, S., Borodkin, A. “Optimal Histogram Equalization Based on Dynamic Programming”, Proceedings of the International Conference on Imaging Science, Systems, and Technology (CISST’02), vol. 1, pp.246-249, CSREA Press, 2002.

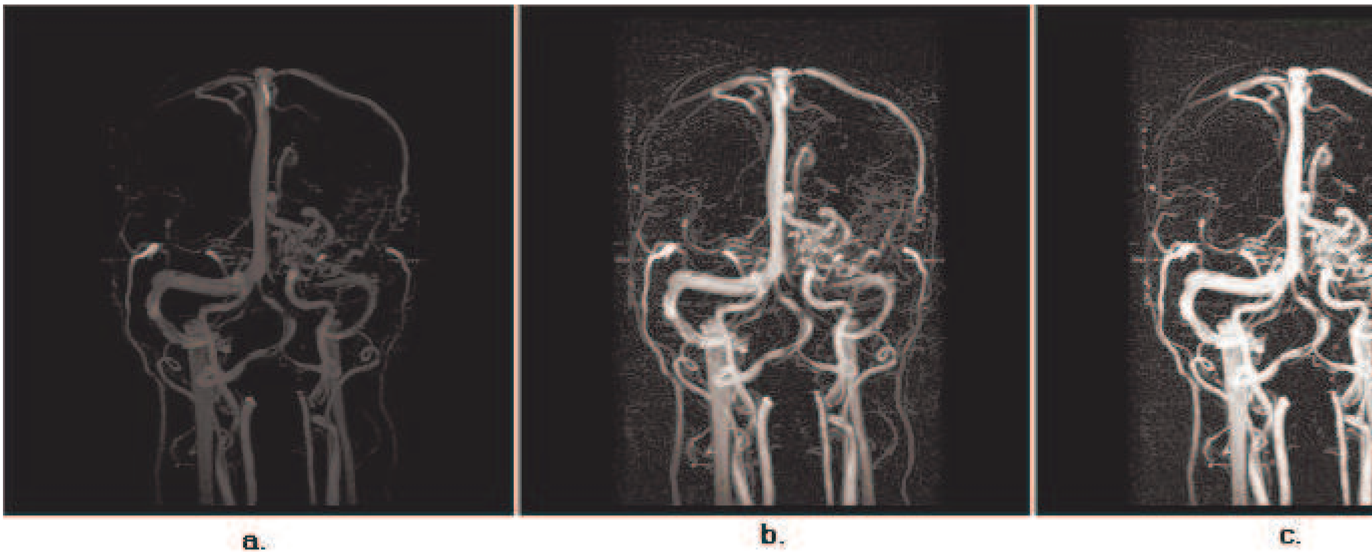

Fig. 1. Image *MR-ANGIO* (Source:

ftp://ftp.erl.wustl.edu/pub/dicom/images/version3/other/philips/mr-angio.dcm.Z)

a. Linear mapping

b. Optimal gray scale reduction

c. Histogram equalization

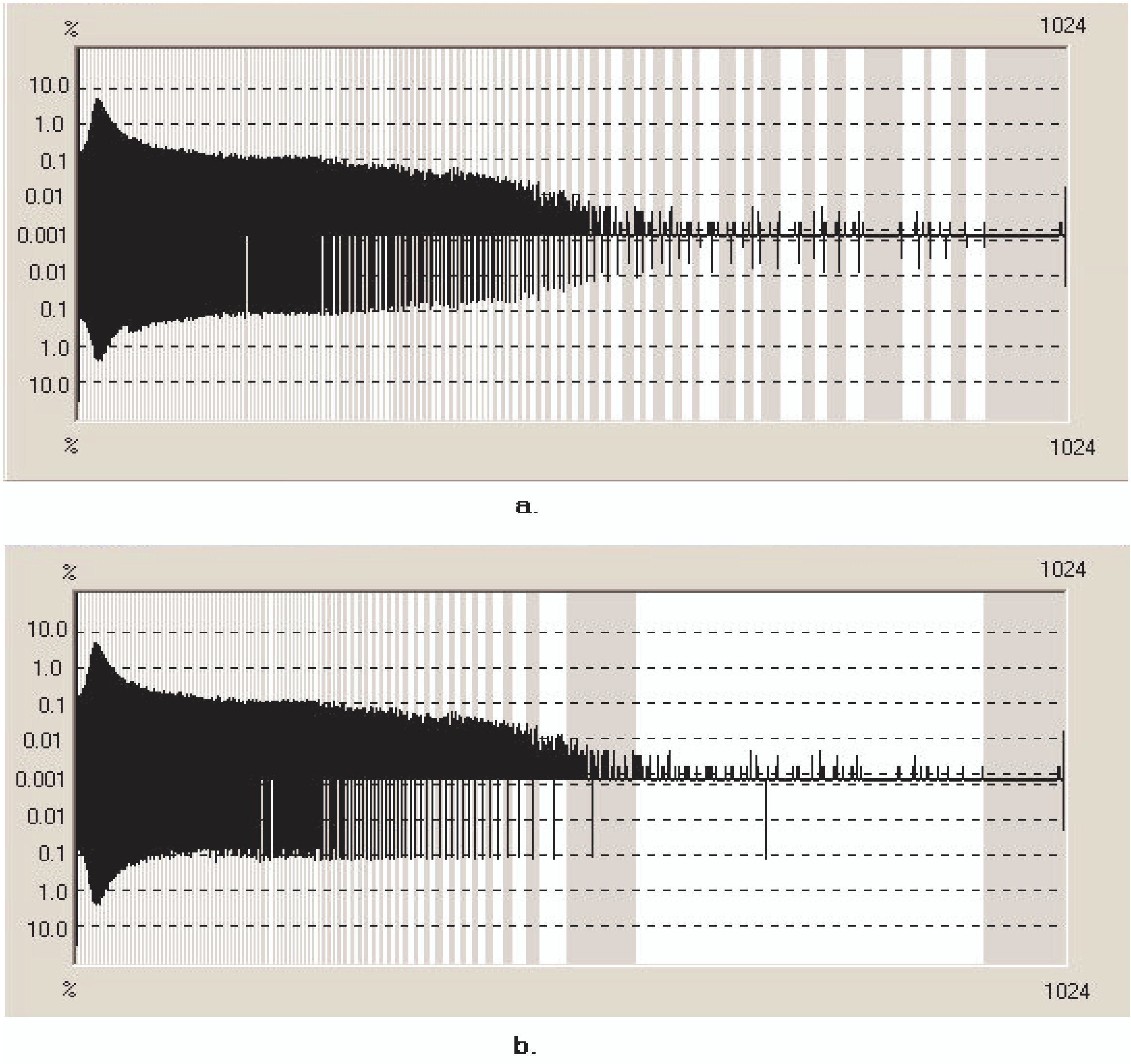

Fig. 2. Intensity histograms for image *MR-ANGIO* from Fig. 1

a. *OGSR* image versus source image; b. *HE* image versus source image

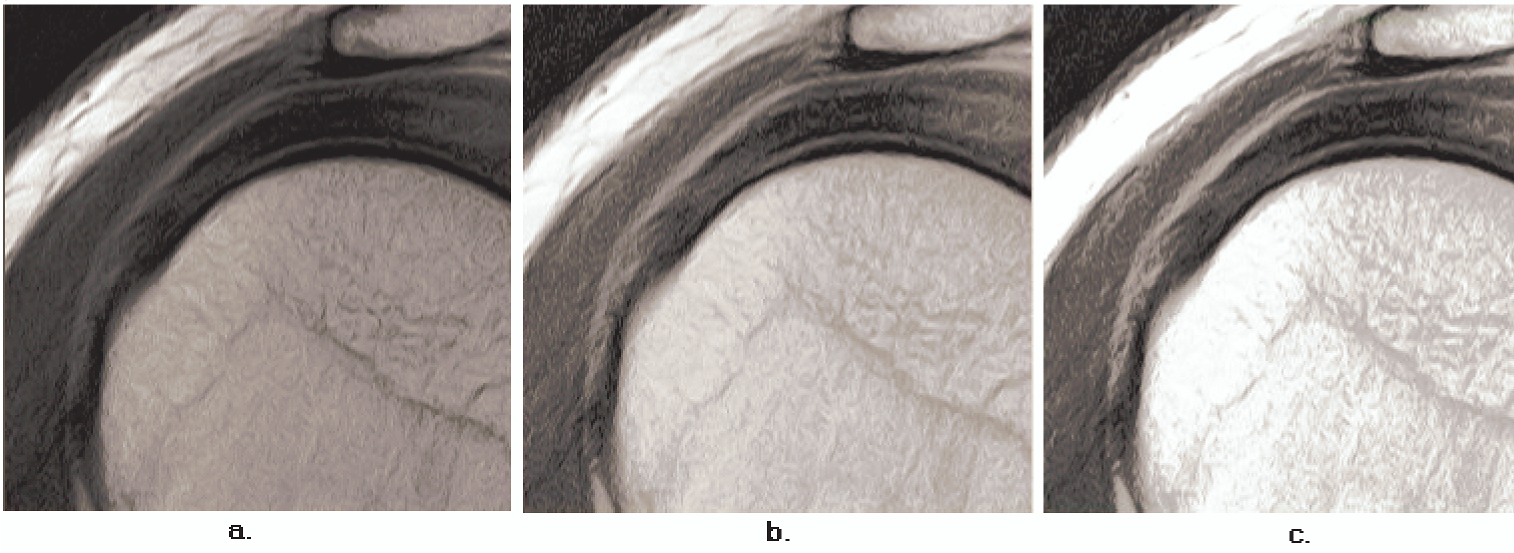

Fig. 3. Image *MR-SHOULDER* (Source:

ftp://ftp.erl.wustl.edu/pub/dicom/images/version3/other/philips/mr-shoulder.dcm.Z)

a. Linear mapping

b. Optimal gray scale reduction

c. Histogram equalization