Statistical comparison of color cancer cell images

EUGENE DEMIDENKO

Dartmouth Medical School, Lebanon, NH 03756, USA

Abstract. A proper analysis and comparison of medical images requires an objective criterion for image discrimination. We developed a statistical approach for color image comparison and showed how it can be applied to discriminate cancer histology images from four treatment groups. We applied a well-established multivariate analysis of variances (MANOVA) on the logit scale so that image comparison was reduced to the classic chi-square test.

Introduction

Visual image evaluation and image comparison is a routine practice of clinicians and pathologists. However, it requires special training and years of practice to acquire the ability to recognize tumorous and benign lesions in sample of images. Specialists sometimes have different opinions on the presence of a tumor, and subjective errors are inevitable. Moreover, evaluation using the eye limits the image comparison to two or three. Today, we are able to obtain dozens or hundreds of images from different samples, locations, at different time periods, etc. The development of objective microscopic image comparison is a burning issue of cancer research and practice. Without denying medical expertise, attempts to develop objective measures for statistical comparison of ensembles of microscopic medical images have been made.

The first attempts at comparing medical images using grayscale distribution are described in a recent book (1). A nonparametric Kolmogorov-Smirnov test is applied to compare images using the cumulative distribution function of gray intensities (2). The goal of the present study is to extend previous methodologies to color cancer cell images. Compared to grayscale images, the use of color image increases statistical power up to the order of three.

Correspondence to: Professor Eugene Demidenko, Dartmouth Medical School, Lebanon, NH 03756, USA
E-mail: eugene.demidenko@dartmouth.edu

Materials and methods

Motivating example. As a motivating example, cancer cell images from four groups of mice with different treatment regimens were considered (5). With six images in each group, only the first two images are shown in Fig. 1 due to the paper length limitation. These 512x384 histology images were acquired under the same illumination and magnification and represent proliferating activity of tumors exposed to EB 1089, radiation, and 1089+radiation by the proliferation marker Ki-67 staining of paraffin-embedded tumor tissue sections. Since the dark spots represent live cancer cells, we can quantify images by color intensity (the darker the image, the more cancer cells). An important feature of these data is that images vary between groups, as well as within groups.

To show that colors correlate, Fig. 2 shows a scatter plot of RGB color components for the first image in the control group (each point on the graph is a pixel with color intensity in the range of 0-255 at the axes). The correlation between colors is high, i.e. the red and green colors correlate with the coefficient of determination, $R^2=0.94$. This means that the variance of green color can be explained by the red color with intensity 94%.

MANOVA for color images. We developed a multivariate analysis of variance (MANOVA) model for color images, which served as the basis for our color image comparison. The MANOVA model is a generalization of the ANOVA model and is a special case of linear model (3,4). The key assumption of our approach is that a microscopic image can be sufficiently characterized by its histogram, so two images are considered to be the same if their histograms are the same. Two images with different content may have the same histogram. Thus, our approach is valid only when images of the same content are compared. Cellular and microscopic biomedical images are generally good examples when the histogram is a satisfactory classifier (1). One could use histogram values to discriminate images, but the log transformation is preferred because histogram values are positive and can be added to 1, so the usual hypothesis testing which requires Gaussian distribution is not valid. Instead of histogram values h_g with g=0-255, we work with logits defined as $y_g=\ln(h_g/h_0)$, where g=1-255 is the gray level and h_0 is the frequency of the background (black color). Since a color image is equivalent to three grayscale images, the logit transformation was applied to each color component (R,G, and B). Logit may be interpreted as the relative color
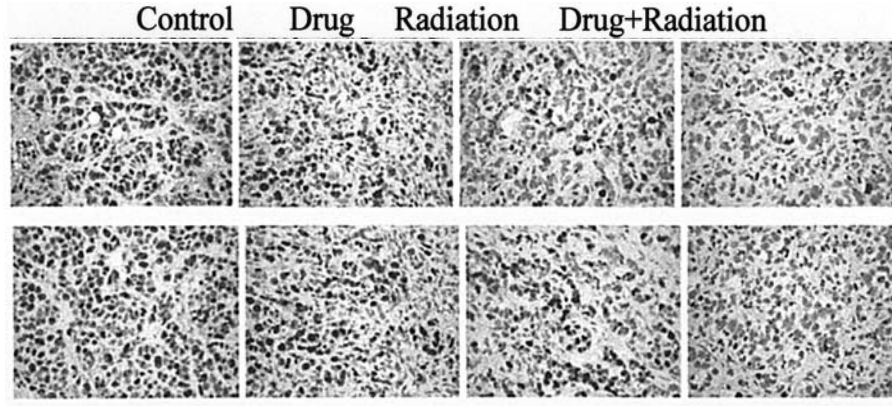
Figure 1. Four treatment groups of cancer images (although there are 6 images per group, only 2 are shown to save space). The dark spots represent cancer cells.
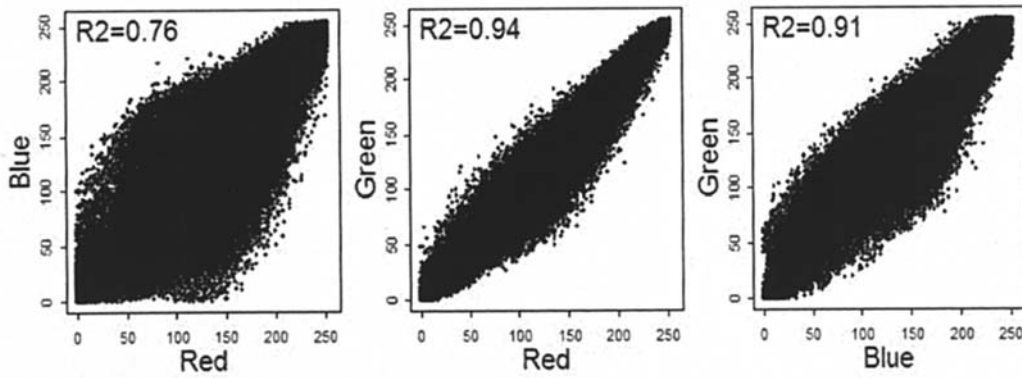

Figure 2. Scatter plot of three colors for the first image in the control group. The correlation between colors is high.

Table I. Standard errors of logits for three color components.

Standard error of logit	Red	Green	Blue
Within-group SE	0.827	0.773	0.840
Between-group SE	0.974	0.914	0.987
Total SE	1.279	1.197	1.297

intensity with respect to background on the log scale. Logits may take negative and positive values and cannot be added to 1, so 255 instead of 256 values were used. Logits for 6x4 images were plotted and are shown in Fig. 3. Note that the pattern for each color is similar, and variation within the group is substantial.

The statistical model for color images in terms of logits was formulated in general terms using n groups of color images with m images in each group. Assuming that images are in RGB format, a 255x3 matrix of logits $\mathbf{y}_{ij}$ is created. (Boldface font is used to indicate vectors and matrices.) The underlying assumption of our model is that images from the same group have the same logit up to an arbitrary shift. Thus, image intensities are allowed to vary within the group, keeping the variation constant across 255 logits. These assumptions lead to the following two-way MANOVA model $\mathbf{y}_{ij}=\boldsymbol{\mu}_i+\mathbf{1}\boldsymbol{\delta}_j^T+\mathbf{e}_{ij}$ where i=1...n and j=1...m, $\boldsymbol{\mu}_i$ is the 255x3 matrix of logits of the i-th group (gross mean), $\boldsymbol{\delta}_j$ is the 3x1 vector of deviations of the individual logit from the average, and $\mathbf{1}$ is the 255x1 vector with unity components (^{T} is the transposition symbol). Parameter vectors $\boldsymbol{\mu}_i$ and $\boldsymbol{\delta}_j$ are unknown and subject to estimation (without loss of generality, one can assume that $\sum \boldsymbol{\delta}_j=\mathbf{0}$). All 255x3 elements of the error matrix $\mathbf{e}_{ij}$ have Gaussian distribution with zero mean and independent columns. However, the three colors are dependent, and the correlation is high as shown in Fig. 2. This means that the elements in each row correlate. Moreover, we assume that each row has the same Gaussian distribution with the 3x3 covariance matrix $\text{cov}(\mathbf{e}_{ij})=\Omega=\text{cov}(e_{ic},e_{ic'})$, where c and c' take values R,G, or B. For example, the (1,1)-th element of this matrix is the variance of the red logit, and $\sigma_{RG}/(\sigma_R\sigma_G)$ is the correlation coefficient between red and green on the logit scale, etc. Parameters of the MANOVA model are estimated by maximum likelihood. Omitting the derivation, the gross mean and individual deviation are estimated as $\mathbf{m}_i=\sum\mathbf{y}_{ij}/m$ and $\mathbf{d}_j=\sum(\mathbf{y}_{ij}-\mathbf{m}_i)^T\mathbf{1}/(255n)$, and the covariance matrix of the colors is estimated as $\Omega=\sum\sum(\mathbf{y}_{ij}-\mathbf{m}_i-\mathbf{1}\mathbf{d}_j^T)^T(\mathbf{y}_{ij}-\mathbf{m}_i-\mathbf{1}\mathbf{d}_j^T)/(256nm)$, where summation is over i=1...n and j=1...m.

The presence of term $\boldsymbol{\delta}_j$ in our model allows intra group variation; particularly, $\sum\delta_j^2/m$ is associated with the within-group variation of color j=c.
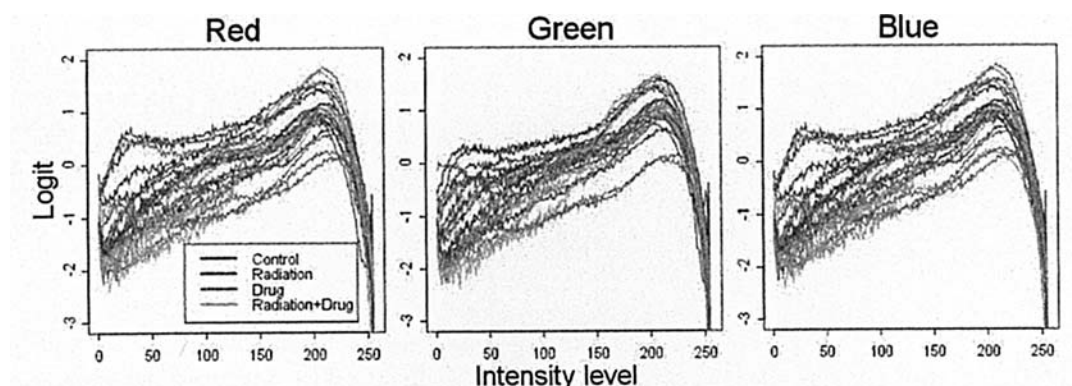
Figure 3. Logit plot for four treatment groups.

Hypothesis testing for image comparison. After parameters of the model are estimated, various hypotheses can be tested, i.e. if one is interested in testing whether color images from group i and j are the same, the null hypothesis is $H_0: \mu_i = \mu_j$. Since the covariance matrix of each row of matrix $\mathbf{m}_i$ is estimated as Ω/m according to the classical chi-square test $(m/2)\text{tr}[(\mu_i - \mu_j)\Omega^{-1}(\mu_i - \mu_j)^T] \sim \chi^2(756)$, where factor 2 is used as images from different groups are independent, $3 \times 255 = 756$ is the degree of freedom. If the left-hand side (trace value) is larger than the critical value of the chi-square distribution, then the hypothesis that images from groups i and j are the same is rejected.

Results

The chi-square test was applied to four cancer cell groups (Fig. 1). With the goal of comparing the cancer cell images in four groups and proving that cancer treatments are statistically different, the estimates $\mathbf{m}_i$ and δ_j (not shown) summarize the analysis of color variances (Table I). In the first row, the standard error of logit within-group variation is computed using the square root of $\sum \delta_{jc}^2/m$. In the second row, the between-group SE is computed using the square root of the diagonal element of matrix Ω, i.e. for red, $\sigma_R = 0.974$. The total variation is the sum of the between- and within-group variations; in other words, the third row is the square root of the sum of squared values of the first and second rows. These values estimate the vertical width of the logit curves in Fig. 3. The variations of red and blue logits are close, while the variation of green is slightly less.

To test the difference between images in four groups, the chi-square test was applied. The critical value of the chi-square distribution with probability p=0.99 (test sensitivity) and 768 d.f. is 862 (the 99th percentile), and the trace value is not less than 2000 for any group comparison. This means that all group images are different with a probability error of 1%, and cancer treatments lead to statistically different cell kill effects.

To the best of our knowledge, this is the first study where an objective method was developed to statistically compare groups of microscopic medical images based on the grayscale intensity distribution for all three colors, unlike traditional eyeball image assessment. An important feature of our approach is that it distinguishes the variation within and between groups. Therefore, it may serve as statistical proof of the treatment effect and be applied to multiple medical image comparison in an automatic regime.

References

1. Demidenko E: Mixed Models: Theory and Applications. John Wiley and Sons, New York, NY, 2004.
2. Demidenko E: Kolmogorov-Smirnov image comparison. Lecture Notes Comp Sci 3056: 933-938, 2004.
3. Rao CR: Linear Statistical Inference and Its Application, 2nd edition. John Wiley, New York, NY, 1973.
4. Searle SR: Linear Models. John Wiley, New York, NY, 1971.
5. Sundaram S, Sea A, Feldman S, Strawbridge R, Hoopes PJ, Demidenko E, Binderup L and Gewirtz DA: The combination of a potent vitamin D3 analog, EB 1089, with ionizing radiation reduces tumor growth and induces apoptosis of MCF-7 breast tumor xenografts in nude mice. Clin Cancer Res 9: 2350-2356, 2003.