A New Metric for Grey-Scale Image Comparison

DALE L. WILSON
Department of Engineering Science, The University of Oxford
dale@robots.ox.ac.uk

ADRIAN J. BADDELEY
Department of Mathematics, The University of Western Australia
adrian@maths.uwa.edu.au

ROBYN A. OWENS
Department of Computer Science, The University of Western Australia
robyn@cs.uwa.edu.au

Received May 5, 1995; Revised February 2, 1996; Accepted January 29, 1997

Abstract. Error measures can be used to numerically assess the differences between two images. Much work has been done on binary error measures, but little on objective metrics for grey-scale images. In our discussion here we introduce a new grey-scale measure, $\Delta_g$, aiming to improve upon the most common grey-scale error measure, the root-mean-square error. Our new measure is an extension of the authors’ recently developed binary error measure, $\Delta_b$, not only in structure, but also having both a theoretical and intuitive basis. We consider the similarities between $\Delta_b$ and $\Delta_g$ when tested in practice on binary images, and present results comparing $\Delta_g$ to the root-mean-squared error and the Sobolev norm for various binary and grey-scale images. There are no previous examples where the last of these measures, the Sobolev norm, has been implemented for this purpose.

Keywords: grey-scale image comparison, error measures, $\Delta$ metrics, root-mean-squared error, Sobolev norm, Hausdorff metric, Myopic topology, image distortion, visual perception

1. Introduction

There are three main methods for comparing images: human perception, which is a subjective measure; objective measures based on theoretical models; and measures defined mathematically, but based on models of the human visual system. In this paper we look at the second class of these methods with regards to objective comparison of grey-scale images, beginning with a brief discussion of the most common of these, the root-mean-squared error, along with the signal-to-noise ratio and the Sobolev norm.

It is widely accepted that currently used binary error measures are those most suited for application on binary images. We provide brief theoretical reasoning in Section 3 for a recently derived binary error measure, $\Delta_b$. This measure, satisfying the axioms of a metric, calculates the “distance” between two sets of object pixels as a means of evaluating the numerical difference between two images. In Section 4 we introduce a similar technique for comparing the “distance” between two grey-scale images, deriving the error metric $\Delta_g$, which is an extension of $\Delta_b$. A grey-scale image can be modelled by a particular type of upper semicontinuous function defined by Serra as a picture function (Serra, 1982), which is a mapping from a subset of $\mathbb{R}^2$ into a subset of $\mathbb{R}$. The “distance” between two picture functions can be defined as the “distance” between the volumes...
beneath the two functions. The metric $\Delta_g$ calculates this distance in either discrete or continuous space.

In the final section of this paper we present the results of testing $\Delta_b$, $\Delta_s$, the root-mean-squared error and the Sobolev norm on various binary and grey-scale images. An original image is compared with several approximations to it, each displaying specific distortions. The discussion focuses on the comparative performance of the four measures and their relationship to the visible differences. Although a variety of distortions are covered in the examination, the types of images used are very restricted and hence the topic is left open for further investigation.

2. Images and Error Measures

Comparison between binary images is usually made numerically using either of two techniques stemming from the study of edge detection algorithms: counting the number of pixels incorrectly detected (how many false positives in response to an edge, and how many missed edges), and measuring the localisation of these errors (how close the response to an edge is) (Baddeley, 1992a; Canny, 1986). Much discussion of the latter has been made with regard to edge detection (Abdou and Pratt, 1979; Baddeley, 1992a, 1992b; Pratt, 1991; Tagare and deFigueiredo, 1990). Although these methods are reasonable measures of error, the rates of error computed using either type of binary error measure are not entirely satisfactory, and their pitfalls have been noted (Baddeley, 1992a, 1992b; Peli and Malah, 1982).

The situation is complicated further with the addition of more possible grey levels, as in grey-scale images. In this case error measures currently used to compare two grey-scale images, such as the root-mean-squared (RMS) error and the root-mean-squared signal-to-noise ratio (SNR) simply incorporate tallies over the two images of corresponding pixel differences, regardless of position or intensity. We give a brief description of these two measures.

Let $f$ and $g$ be two grey-scale images. Then the root-mean-squared error is given by

$$\text{RMS}(f, g) = \left[ \frac{1}{n(X)} \sum_{x \in X} (f(x) - g(x))^2 \right]^{1/2},$$

where $n(X)$ is the number of pixels in the pixel raster $X$, and the root-mean-squared signal-to-noise ratio is defined by

$$\text{SNR}(f, g) = \left[ \frac{\sum_{x \in X} g(x)^2}{\sum_{x \in X} (f(x) - g(x))^2} \right]^{1/2}.$$

A criticism made of the RMS error and other similar error measures is that they are inaccurate predictors of perceptual distortion (Teo and Heeger, 1994a, 1994b; Watson, 1993). One factor compounding this inaccuracy is that these measures have the effect that errors of equal magnitude have the same visual impact regardless of their intensities. For human perception this is not the case, since the retina has different sensitivities to different intensities of light. Daly (1993) proposes a human visual system model, whereby an error measure incorporates a nonlinear sensitivity function. The most likely choice is a logarithmic function (Cornsweet, 1970), so that in practice each pixel value in both $f$ and $g$ could be transformed using some sensitivity function, and the RMS error or SNR calculated from the new values.

Despite its apparent faults the RMS error has many theoretical benefits and is very straightforward to calculate, giving it computational advantages. As a consequence of its simplicity and generality examples of the use of RMS are frequently found in the development of image compression algorithms, their parameters often selected to minimise RMS error (Delp and Mitchell, 1979; Gonzalez and Woods, 1993; Habibi, 1974; Jain, 1981; Linde, 1980; Netravali and Haskell, 1988; Rabbani and Jones, 1991; Woods and O’Neil, 1986).

Another possibility for grey-scale image comparison is in using the Fourier transform of images. For an image $f$, the discrete Fourier transform can be given by

$$F(u) = \sum_{x \in X} f(x) \exp\left( \frac{-i2\pi u x}{N} \right),$$

for each pixel $u$ in the frequency domain $U$ (Bracewell, 1986). The link, with respect to error measures, between the spatial representation of an image and the frequency domain representation can be shown using Parseval’s Identity:

$$\sum_{x \in X} |f(x)|^2 = \frac{1}{n(U)} \left[ \sum_{u \in U} |F(u)|^2 \right].$$

From this we can derive an equivalent version of the RMS error in the frequency domain,

$$\left[ \frac{1}{n(X)} \sum_{x \in X} |f(x) - g(x)|^2 \right]^{1/2} = \left[ \frac{1}{(n(U))^2} \sum_{u \in U} |F(u) - G(u)|^2 \right]^{1/2}, \quad (2.1)$$

where $F$ and $G$ are the discrete Fourier transforms of $f$ and $g$ respectively. See Bracewell (1986) for conditions.