Neurocomputing 137 (2014) 65–70

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Image matching using moment invariants

Prashan Premaratne [a,*], Malin Premaratne [b]

[a] School of Electrical Computer and Telecommunications Engineering, University of Wollongong, North Wollongong, NSW 2522, Australia
[b] Department of Electrical and Computer Systems Engineering at Monash University, Australia

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form
23 February 2013
Accepted 28 February 2013
Available online 17 February 2014

Keywords:
Image similarity
Structural similarity
Moment invariants
SSIM
MISM

ABSTRACT

Matching images using Mean Squared Error (MSE) and Peak Signal to Noise (PSNR) ratios does not well conform to the Human Visual System (HVS). When matching two images, HVS operates both globally and locally when it identifies features of a scenery and this process is not matched adequately by PSNR or MSE. A low MSE or very high PSNR may not necessarily mean that images are similar. Similarly, when images are similar as HVS would identify, the corresponding MSE may not be very low and PSNR may not be very high. However, quite recently, a new measure has been proposed to circumvent the drawbacks of PSNR or MSE. This measure known as Structural Similarity Measure (SSIM) has received acclaim due to its ability to produce results on a par with Human Visual System. However, experimental results indicate that noise and blur seriously degrade the performance of the SSIM metric. Furthermore, despite SSIM's popularity, it does not provide adequate insight into how it handles 'structural similarity' of images. We propose a new structural similarity measure based on approximation level of a given discrete Wavelet decomposition that evaluates moment invariants to capture the structural similarity with superior results over SSIM.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Today, there are thousands of image databases in use by different countries to assist in law enforcement in analyzing stolen art, look for counterfeit notes, face detection [1,2] at entry points and many other scientific analysis. Comparing a section of a counterfeit note against a sample with high precision would be a tremendous advantage for the law enforcement authorities. Then there are human faces that differ due to aging, illness or harsh environmental conditions may escape the scrutiny of the law enforcement due to weaknesses in algorithms that might not be as flexible as human perception. Striking a balance between these two extreme ends and modeling such requirements into a metric would not be a small task. Every day, a countless number of images are uploaded onto Internet and other databases by millions of users as part of their occupation or pastime. Law enforcement agencies such as FBI in the USA have successfully developed the worlds largest Fingerprint database which is used extremely successfully for the past decade. Gatwick airports at London in the UK have successfully used iris detectors to check immigrant for illegal entry over the past 7 years. These are some of the image matching approaches being successfully run around the world mimicking human ability to detect with machine precision. Some of the above success stories have been heard due to the nature of the controlled environments that they are being operated in. An iris scanner would prompt a user to pose for a capturing device in an orderly manner. A finger print will be nothing more than few lines and they can be enhanced if not properly captured. However, being able to run through online items for sale looking for a stolen artefact would be much more challenging as these uploaded images are taken from different angles under different lighting conditions. Such uncontrolled environments pose a great challenge in identifying faces when people walk in crowds. Humans are quite capable of identifying faces even under disguises. Yet, face recognition is a monumental task in uncontrolled environments.

Earlier attempts to match images relied on single discriminants such as color histograms [3]. However, different images may have very similar histograms and when large databases are searched, single discriminant approach does have many drawbacks. Another well-known approach has been to Mutual Information (MI). This refers to the concept of finding the relationship between corresponding individual pixels disregarding the pixel's respective neighborhood [4]. This results in loosing much of the spatial information further distancing it from the Human Visual System. Modern techniques rely on few features extracted from images such as color, texture and shape [5–7]. Different approaches have developed different metrics to directly compare images using compression coefficients such as the Karhueneen–Loeve transformation [8]. However, researchers are unsure whether

* Corresponding author.
E-mail address: prashan@uow.edu.au (P. Premaratne).

http://dx.doi.org/10.1016/j.neucom.2013.02.058
0925-2312 © 2014 Elsevier B.V. All rights reserved.

these metrics have a good correlation with the human perception of image similarity [9].

2. Image similarity basics

2.1. Pixel based image similarity

When comparing two images, the most obvious approach is to match corresponding pixels. This is the simplest approach in which two corresponding pixels of images are compared giving rise to the following expression:

$$Pd1 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}(f(i,j)-g(i,j))^2}. \quad (1)$$

Here the two images f and g are of sizes $m \times n$. Another metric based on pixel difference can be calculated as

$$Pd2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(f(i,j)-g(i,j))^2 \quad (2)$$

and is known as the Mean Squared Error (MSE) and has been used extensively over the decades. The main drawback of MSE is that it does not correlate well with human perception. Image similarity can also be assessed using correlation. The following equation denotes the normalized correlation where correlation between two images has been divided by square rooted autocorrelation of both target and reference objects:

$$Pd3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} f(i,j) * g(i,j)}{\sqrt{\sum_{i=1}^{n} f(i,j)^2 \sum_{j=1}^{m} g(i,j)^2}}. \quad (3)$$

Another interesting metric namely Pattern Intensity (PI) operates on the assumption that the structures would vanish from the difference if the images are aligned [10]. A pixel is considered to belong to a structure if it has a significantly different intensity from its neighboring pixels. For two images f and g the following expressions show the PI:

$$I(i,j) = f(i,j) - kg^T(i,j)$$
$$PI_{r,\sigma} = \sum_{i,j}\sum_{d^2 \le \sigma^2} \frac{\sigma^2}{\sigma^2 + ((I(i,j)-I(v,w))^2)}.$$
$$d^2 = (i-v)^2 + (j-w)^2. \quad (4)$$

As stated above, all the pixels within a radius r are considered in the calculation. The role of σ is to weight the function so that small variations in intensity caused by noise due to varying lighting conditions would not result in significant changes in metric [4]. These different metrics do not provide any comprehensive solution for image matching that would agree with HVS. Many researchers have felt that a suitable combination of metrics would provide an amicable solution.

2.2. Combined metrics

Comparing two images accurately to ascertain whether there is a match or not is essential for many image processing related applications such as watermarking, compression and content retrieval. As pointed out before, age-old metrics such as Mean Squared Error (MSE) have been used for decades despite its inability to agree with human subjective analysis [11,12]. Recently, light has been shed on a new metric that seems to agree with Human Visual System [12]. The SSIM is supposed to estimate image degradation as perceived change in structural information. Structural information contains strong inter-dependencies of pixels especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. SSIM has been singled out due to its claim of superiority over the existing metrics [13,14]. However, it has been observed that SSIM does not perform well with blurred images [13]. Since a blurred version of an image essentially contains the same structure, SSIM's inability to measure the structural similarity of blurred images raises an issue as to whether SSIM does truly look for the structural content. From our research, we have concluded that despite SSIM's claim of superiority, its ability to compare similar structures is doubtful as will be demonstrated in Section 4. We have developed a new metric that uses some of the concepts exploited by SSIM [15]. The new metric demonstrates better performance over SSIM in blurred images and images corrupted by Gaussian and Salt & Pepper noise.

2.3. Structural similarity measure

SSIM attempts to separate the task of similarity measurement of two images into luminance, contrast and structure [13]. Hence, a similarity measure is defined as

$$SSIM(\mathbf{P_1}, \mathbf{P_2}) = l(\mathbf{P_1}, \mathbf{P_2}) \times c(\mathbf{P_1}, \mathbf{P_2}) \times s(\mathbf{P_1}, \mathbf{P_2}) \quad (5)$$

where $\mathbf{P_1}$ and $\mathbf{P_2}$ are the two images being compared and l, c and s stand for luminosity, contrast and similarity measures respectively. Mean μ and standard deviation σ of images $\mathbf{P_1}$ and $\mathbf{P_2}$ are defined as follows:

$$\mu_{P1} = \frac{1}{M \times N}\sum_{y=0}^{M-1}\sum_{x=0}^{N-1}\mathbf{P_1}(x,y)$$

$$\sigma_{P1} = \sqrt{\frac{1}{(M \times N-1)}\sum_{y=0}^{M-1}\sum_{x=0}^{N-1}(\mathbf{P_1}(x,y)-\mu_{P1})^2}$$

$$\sigma_{P1}\sigma_{P2} = \frac{1}{M \times N-1}\sum_{y=0}^{M-1}\sum_{x=0}^{N-1}(\mathbf{P}_1(x,y)-\mu_{P1})(\mathbf{P}_2(x,y)-\mu_{P2})$$

Furthermore, l, c and s for images $\mathbf{P_1}$ and $\mathbf{P_2}$ are calculated as follows [13]:

$$l(\mathbf{P_1}, \mathbf{P_2}) = \frac{2\mu_{P1}\mu_{P2} + C_1}{\mu_{P1}^2 + \mu_{P2}^2 C_1}$$

$$c(\mathbf{P}_1, \mathbf{P}_2) = \frac{2\sigma_{P1}\sigma_{P2} + C_2}{\sigma_{P1}^2 + \sigma_{P2}^2 + C_2}$$

$$s(\mathbf{P}_1, \mathbf{P}_2) = \frac{\sigma_{P1P2} + C_3}{\sigma_{P1}\sigma_{P2} + C_3}.$$

Constants C_1, C_2 and C_3 are used for the stability of equations when they are extremely small. Combining the above definitions, Eq. (1) can be expressed as follows when $C_3 = C_2/2$ for simplicity [13]:

$$SSIM(\mathbf{P}_1, \mathbf{P}_2) = \frac{(2\mu_{P1}\mu_{P2} + C_1)(\sigma_{P1P2} + C_2)}{(\mu_{P1}^2 + \mu_{P2}^2 C_1)(\sigma_{P1}^2 + \sigma_{P2}^2 + C_2)}.$$

The expression $s(\mathbf{P}_1, \mathbf{P}_2)$ is a simple function of cross correlation. It does not contain any notion of structure as structure in an image would represent directionality of objects and how they are organized. Cross correlation would only capture similarity of pixels and not structure. This clearly indicates that SSIM does not evaluate structure and hence should not be misrepresented as evaluating structure. Section 4 would clearly indicate the evidence of SSIM's inability to evaluate structure.

Download English Version:

https://daneshyari.com/en/article/409923

Download Persian Version:

https://daneshyari.com/article/409923

Daneshyari.com