Signal Processing: *Image Communication* 31 (2015) 61–75

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

An oriented gradient based image quality metric for pedestrian detection performance evaluation

Anthony Winterlich [a,*], Ciarán Hughes [b], Liam Kilmartin [b], Martin Glavin [b], Edward Jones [b]

[a] *Department of Electrical and Electronic Engineering, NUI, Galway, Ireland*
[b] *Valeo Vision Systems, Dunmore Road, Tuam, Co. Galway, Ireland*

ARTICLE INFO

Article history:
Received 20 July 2014
Received in revised form
6 December 2014
Accepted 6 December 2014
Available online 17 December 2014

Keywords:
Image quality
Automotive machine vision
Objective quality
Histogram of oriented gradients

ABSTRACT

Full reference image quality metrics are important tools for optimizing system design parameters associated with image acquisition, compression and transmission. While optimizing systems for perceptual quality is important, in the automotive environment Advanced Driver Assistance Systems (ADAS) such as automated pedestrian detection are becoming a common feature of in-vehicular vision systems. As such, automotive image quality must also be tuned for optimal machine vision performance.

In this paper the effects of transmission artifacts on the performance of a number of state-of-the-art pedestrian detection algorithms are evaluated. We demonstrate that the human visual system may not perceive distortions that adversely affect machine vision performance. As a result, existing full-reference image quality metrics are not necessarily accurate predictors of machine vision performance on transmitted video sequences. To address this problem, a novel, computationally inexpensive, full-reference objective quality metric based on histogram of oriented gradients is proposed. The proposed metric accurately predicts algorithm performance in the presence of transmission artifacts. The metric can be used at the system design stage in order to optimize image capture parameters for machine vision performance without the need for annotated test databases, which are both expensive and time consuming to produce.

1. Introduction

Machine vision functionality is becoming increasingly important in automotive systems [1]. An important class of applications is the rapid detection of vulnerable road users such as pedestrians, or other vehicles that pose a threat to the host vehicle. In particular, the task of pedestrian detection from cameras mounted on a vehicle has become increasingly important to a number of automotive safety applications such as collision avoidance and autonomous driving [2–6].

In typical automotive vision systems degradation in image quality can occur as a result of lossy compression or transmission errors. Quantifying the loss of performance of detection algorithms due to these degradations typically

Abbreviations: ADAS, Advanced Driver Assistance Systems; AP, Average Precision; AWGN, Additive White Gaussian Noise; CR, Compression Ratio; DCT, Discrete Cosine Transfer; DPM, Deformable Parts Model; FPPI, False positives Per Image; FSIM, Feature Similarity Index; GM, Gradient Magnitude; GSM, Gaussian Scale Mixture; HMSE, Hog Mean Squared Error; HOG, Histogram of Oriented Gradients; HVS, Human Visual System; ICF, Integral Channel Features; IFC, Information Fidelity Criterion; INRIA, *Institut National de Recherche en Informatique et en Automatique*; IQA, Image Quality Assessment; JP2, JPEG2000; JPEG, Joint Photographic Experts Group; IWSSIM, Information-content Weighted Structural Similarity; L2-Hys, Lowe-style clipped L2 norm; LAMR, Log-Average Miss Rate; MOS, Mean Opinion Score; MSE, Mean Squared Error; NQM, Noise Quality Measure; PC, Phase Congruency; PLCC, Pearson's Linear Correlation Coefficient; PSNR, Peak Signal to Noise Ratio; RMSE, Root Mean Squared Error; SIFT, Scale Invariant Feature Transform; SR-SIM, Spectral Residual Similarity; SSIM, Structural Similarity; SVM, Support Vector Machine; VIF, Visual Information Fidelity; VSNR, Visual Signal to Noise Ratio

* Corresponding author. Tel.: +353 872192507.
E-mail address: anthonywinterlich@gmail.com (A. Winterlich).

http://dx.doi.org/10.1016/j.image.2014.12.001
0923-5965/

requires the use of annotated test databases, which are both expensive and time consuming to produce. It would therefore be desirable to utilize an objective Image Quality Assessment (IQA) metric that accurately predicts the performance of machine vision algorithms on degraded images and video. Such a metric could be used as a predictor of the impact of choice of system parameters (e.g. Compression Ratio (CR) in image/video compression) on pedestrian detection performance. It may also obviate, or at least reduce the need to carry out extensive machine vision performance evaluation, which requires large annotated datasets.

Video quality is extremely important in automotive vision systems since images displayed to the driver are safety critical. However, the concept of quality is poorly defined in this environment [7]. In [8], the authors make a distinction between the “naturalness” and “usefulness” of an automotive image. The “naturalness” of an image is often referred to as its “perceptual quality”. In order to quantify perceptual quality subjective tests must be carried out, in which a large number of human observers are shown a series of images whose quality they are asked to rate on a particular scale. The mean score of each sequence is termed the Mean Opinion Score (MOS) and is representative of the perceived quality of that image. Such large scale subjective tests have been carried out and have been made publically available to the research community [9–11].

In an automotive context, a natural image should be a faithful representation of the road ahead. For example, it should contain recognizable signal colors and be free from noise and compression artifacts. On the other hand, the same image with exaggerated local contrast and sharpness may be more “useful” if it allows the driver to see more detail, such as a pedestrian in low light conditions. In the automotive environment, the amount of information we can extract from a scene determines its “usefulness”. While these distinctions are important, it is necessary to present a further distinction. Increasingly, image processing algorithms are making use of automotive cameras for applications such as automatic pedestrian or vehicle detection. The question then arises as to whether an image that is useful to the human driver is equally useful to a machine vision algorithm. Unfortunately, this is often not the case, for example, depending on the content of an image, the addition of noise may go unnoticed to a human viewer due to masking effects, however even a modest amount of noise will typically impair the performance of image processing algorithms (see Fig. 1). Similarly, interpolation artifacts may occur at high frequencies that are difficult for a human viewer to discern (as illustrated in Fig. 2), nevertheless the presence of such artifacts are sufficient to degrade the performance of a pedestrian detection algorithm. These examples raise important safety concerns for automotive vision engineers since a seemingly imperceptible change

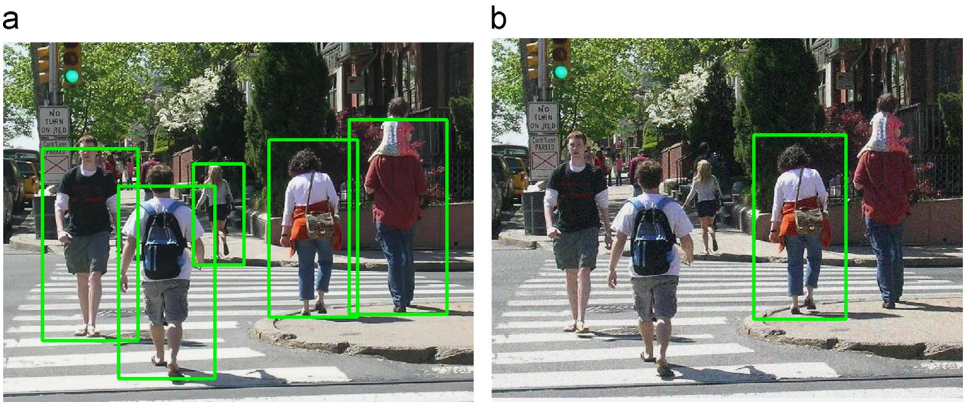

Fig. 1. The presence of noise in (b) can go unnoticed to a human viewer due to masking effects, whereas noise can have a significant effect on the performance of a pedestrian detection algorithm (HOG+SVM).

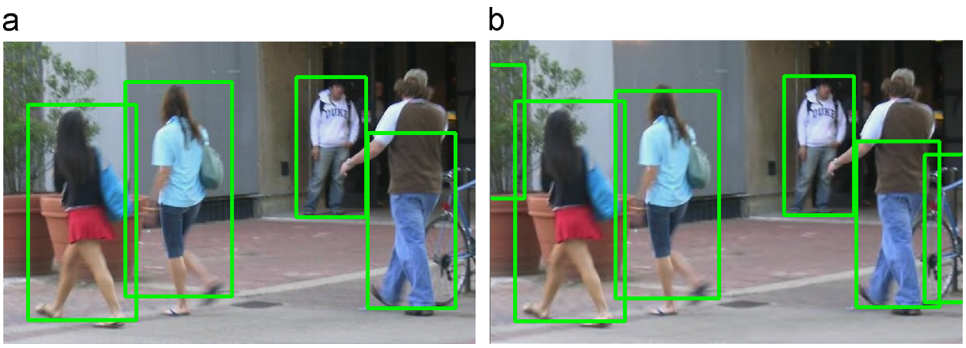

Fig. 2. “Ringing” caused by JP2 compression in (b) can occur at high spatial frequencies that are difficult for the human visual system to discern, nonetheless, the presence of this distortion is sufficient to degrade the performance of a pedestrian detection algorithm (HOG+SVM) by causing an increase in false positives.