ANALYSIS OF PATCH-BASED SIMILARITY METRICS: APPLICATION TO DENOISING

Mounira Ebdelli, Olivier Le Meur**, Christine Guillemot**

INRIA*, University of Rennes 1**
Campus Universitaire de Beaulieu
35042 Rennes Cedex France

ABSTRACT

This paper presents a performance analysis of measures used for assessing similarities between patches. Compared to subjective ground thruth, our results indicate that some metrics are more suitable than others in a context of patch matching. This conclusion is confirmed by an experiment on non-local means (NLM) denoising algorithm. The denoising quality depends on the chosen similarity metric. In the best case, the gain is of 1.3dB compared to a classical SSD-based denoising algorithm.

***Index Terms*—** similarity metric, examplar-based, denoising.

1. INTRODUCTION

The goal of a similarity metric, also called fidelity measure, is to provide a quantitative score that describes the degree of similarity or, conversely, the dissimilarities between two signals [1, 2]. This kind of metric is widely used in image processing algorithms based on patches similarity, so called examplar-based [3]. These algorithms assume that any patch in the image can be approximated with one or more patches in the same image. For example, the denoising algorithm based on NLM updates the pixels values of each patch with a linear combination of the corresponding pixel values in its most similar patches. Thereby, the result of this algorithm is highly dependent on the fidelity metric used to find similar patches. For this purpose, different metrics can be used and can be classified into three main categories : pixel-based, statistic-based and texture-based similarity metrics. Pixel-based metrics compute the similarity between two patches/images using the differences between their pixel values. For instance, the L_p-norm belongs to this category. Statistic-based metrics are based on the similarity between probability density distributions of pixel values in patches. A review of classical statistic-based metrics can be found in [4]. In this paper, we consider the Bhattacharyya probability density [5, 6], the Normalized Mutual Information (NMI) and the Kullback-Leibler Divergence (K-Div). Others texture characteristics-based similarity metrics such as the magnitude of the dominant gradient in patches are proposed in [3, 7, 8].

In this paper we evaluate the performance of a set of similarity metrics. For this purpose, we first perform a subjective analysis in order to compare the similarity results provided by observers to the objective quality scores of the considered fidelity measures. The influence of these metrics is illustrated in the context of denoising with NLM algorithm.

The paper is organized as follows. In Section 2, the classical similarity metrics are presented. A subjective statistical analysis of these metrics is discussed in Section 3. Their performances in the context of denoising are presented in Section 4. Finally, Section 5 concludes the paper.

2. SIMILARITY METRICS: BACKGROUND

This section presents an overview of the most widely used similarity metrics in image processing algorithms. All these metrics provide a similarity score between two candidates. In the following, we present six state-of-the-art metrics belonging to the three aforementioned categories.

2.1. Pixel-based similarity metrics

2.1.1. L_p-norm

The L_p-norm or p-norm of a given signal $x = (x_1, x_2, \ldots, x_n) \in R^n$, $p \geq 1$, is defined as : $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$.

In practice, three main values of p correspond to the three most used norms :

- L_1-norm, or Manhattan metric: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.
- L_2-norm, or Euclidean norm : $\|x\|_2 = (\sum_{i=1}^{n} |x_i|^2)^{\frac{1}{2}}$.
- L_∞-norm, or Chebyshev norm : $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$.

The L_2-norm is the simplest and most used fidelity metric in the literature of image processing. It computes the similarity of two given patches x and y as the sum of squared intensity differences (SSD) of their pixels values: $SSD(x, y) = \sum_{i=1}^{n} (x_i - y_i)^2$. The MSE, which is the mean of SSD, is usually converted into a peak signal-to-noise ratio (PSNR) measure for assessing degradation quality of a noisy image compared to its original version [1]: $PSNR = 10 log_{10} \frac{L^2}{MSE}$. L is the range of pixel intensity (255 for gray-scale image).

This metric has many advantages : symmetry, convexity

and energy preserving property. However, it does not match the perceived visual quality. Two patches x and y having the same degree of similarities with a given patch z i.e. $SSD(x,z) = SSD(y,z)$, may be visually dramatically different.

2.1.2. SSIM (Structural Similarity Information Measure)

Structural Similarity Information Measure (SSIM) computes the similarity between two images or patches x and y. The similarity score is computed based on three terms: luminance ($l(x,y)$), contrast ($c(x,y)$) and structure ($s(x,y)$) as : $SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}$. Here α, β and γ are parameters used to adjust the relative importance of each of the three components [9]. A specific form of the SSIM index where $\alpha = \beta = \gamma = 1$ is usually used:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2\mu_y^2 + C_1)(\sigma_x^2\sigma_y^2 + C_2)}$$

where $C_1 = (K_1 L)^2$, $K_1 << 1$, $C_2 = (K_2 L)^2$, $K_2 << 1$ and $L = 255$ for gray scale images.

2.2. Probabilistic-based similarity metrics

Another approach to compute the similarity between two patches is to consider their probability density function (pdf). This kind of metric provides a high degree of similarity when their pdfs are close. A histogram, which represents the frequency of each interval of values (also called bin), provides the basis for a non-parametric estimation of the pdf [4]. The pdf of a patch x is given by : $p(x) = \frac{h(x)}{n}$, where $h(x)$ is the histogram of the pixel values and n the number of pixels in the patch.

2.2.1. NMI (Normalized Mutual Information)

The mutual information $I(X;Y)$ of two discrete random variables X and Y computes the amount of common information between them as :
$I(X;Y) = \sum_{x_i \in X} \sum_{y_i \in Y} p(x_i, y_i) \log(\frac{p(x_i,y_i)}{p(x_i)p(y_i)})$, where $p(x_i)$ is the probability of x_i. Usually, the normalized form of the mutual information is used : $NMI(X,Y) = \frac{I(X;Y)}{\sqrt{H(X) \times H(Y)}}$. $H(X)$ and $H(Y)$ are respectively the entropy of X and Y defined as : $H(X) = -\sum_{x_i \in X} p(x_i) \log p(x_i)$. NMI metric is bounded between 0 (if X and Y are independent) and 1 (if their pdf are exactly the same).

2.2.2. KL-Div (Kullback-Leibler Divergence)

The Kullback-Leibler divergence (KL-Div) of two probability distributions X and Y is defined as:
$D_{KL}(X \,||\, Y) = H(X,Y) - H(X)$, where
$H(X,Y) = -\sum_{x_i \in X, y_i \in Y} p(x_i, y_i) \log p(x_i, y_i)$.
This metric is non-symmetric and is not upper-bounded. The KL-Div is null when the distributions of the two patches are exactly the same.

2.3. Hybrid distances : Bhattacharyya probability density-based metric

Using NMI or KL-Div metrics to measure similarities between two samples, means that their closeness are only dependant on their probability distributions. Figure 1 shows that this information is not enough. The two patches are different although their probability distributions is the same. To deal with this drawback, the authors in [5] combine the probability distribution with SSD as follow :

$$d(x,y) = \frac{1}{|x|} \times SSD(x,y) \times d_B(x,y) \qquad (1)$$

where $d_B(x,y) = \sqrt{1 - \sum_{i=1}^{B} \sqrt{\rho_x(i) \times \rho_y(i)}}$ is the Bhattacharyya coefficient. ρ_x and ρ_y are the pdf of patches x and y respectively. However, if the two patches have the same distributions the Bhattacharyya coefficient d_B is null no matter for the rotation. To be able to distinguish this case, the distance is computed using the following formula [6]:

$$d(x,y) = \frac{1}{|x|} \times SSD(x,y) \times (1 + d_B(x,y)) \qquad (2)$$

In the rest of the paper we notice Bh_1 the metric distance in equation 1 and Bh_2 the metric in equation 2.

(x) (y)

Fig. 1: Patches of 21×21 pixels estimated as similar using the probability-based metrics ($d_{KL-Div}(x,y) = 0$ and $d_{NMI}(x,y) = 1$) and quite different using SSD ($d_{SSD}(x,y) = 5355$).

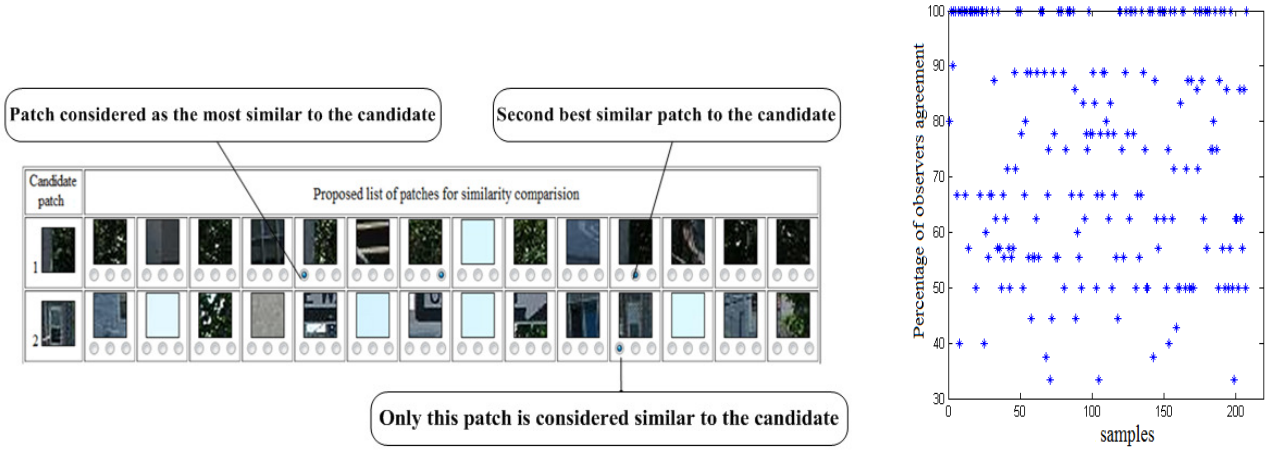
2.4. Dominant orientation Template

The dominant orientation template (DoT) approach is related to the Histograms-of-Gradients (HoG) based representation [7, 8]. It computes the similarity of two patches by comparing the orientations of their dominant gradients. Although this measure does not take into account the difference of texture values, it has the advantage to be robust to noise and illumination change.

3. SUBJECTIVE EXPERIMENT

3.1. Protocol

In order to provide a subjective performance analysis of the aforedescribed similarity metrics, our experiments are conducted on 100 natural images. For each image, two patches of 21×21 pixels are chosen. As shown in figure 2.a, each candidate patch is compared to a list of 14 patches composed with its eight neighbors and six patches randomly selected within the same image. These patches have been presented,



(a) (b)

Fig. 2: (a) Part of the web page used for gathering users similarity decisions; (b) Agreement of observers' similarity decisions.

in a web page [1] to 12 observers who have selected the three first closest patches from the proposed list.

3.2. Results

Once the ranking has been performed, we investigate the subjective and objective similarity scores how they match. For this purpose, we compare the order of the best similar patches given by observers with those obtained by metrics. The general idea is to count the number of matches. For each metric m, we count the number of matches $Corr(m) = \frac{\sum_{i=1}^{N} \delta(i)}{N}$, where N is the total number of patches candidates, and $\delta(i) = \begin{cases} 1 & \text{if } Sim(i,m) = BM(i) \\ 0 & \text{otherwise.} \end{cases}$.
$Sim(i,m)$ gives the index of the best similar patch to the candidate Ψ_i according to the considered metric m whereas $BM(i)$ is the index given by observers.
However, for each candidate patch, the set of proposed patches may contain similar patches, so that observers can not easily choose between them. In this case, observers' scores could be distributed on more than one patch as illustrated in figure 2.b. To deal with this problem, three main constraints for computing the correlation between subjective and objective scores are considered:

- constraint 1: we consider all the subjective scores. A good match between subjective and objective scores is obtained when one of the three most similar patches according to observers corresponds to the best matching patch with respect to the similarity metric.
- constraint 2: same as constraint 1 except that only candidate patches for which more than 50% of the observers have the same decision are considered.
- constraint 3: we consider only candidate patches for which more than 50% of the observers have the same decision and this patch corresponds to the best matching patch with respect to the similarity metric.

Figure 3 illustrates the results of correlation. One can mention that, using each of the considered correlation constraint, the similarity metrics based only on the probability distribution i.e. NMI and KL-Div, have the worst results. This conclusion is expected because the pdf helps to detect the patch structure but do not quantify dissimilarity between the collocated pixels values in the two patches. Metrics based on dominant gradient does not perform well simply because two patches having similar structures may have totally different textures. The best results are given when the SSD and pdf (Bh_1 and Bh_2) are combined.

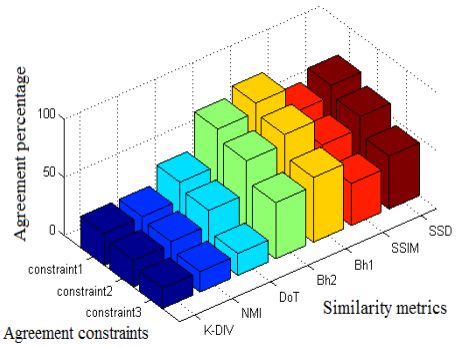


Fig. 3: Agreement between subjective and objective similarity scores.

[1] The web page is reachable at `http://people.irisa.fr/Olivier.Le_Meur/test/test.php`

4. APPLICATION : NLM DENOISING ALGORITHM

In this section we evaluate the influence of the four best correlated metrics with subjective results, i.e. SSD, SSIM, Bh_1 and Bh_2 on an examplar-based denoising application. Each aforementioned metric is used to find the best similar patches within the image. The best patches are then linearly combined by using non local means (NLM) filter. The NLM denoising approach [10][2] considers that each pixel p in the image I can be estimated using a weighted average of collocated pixels values in the most similar patches to the one centered on p (Ψ_p) as : $\hat{I}(p) = \sum_{q \in I} \omega(p,q) I(q)$, where $0 \leq \omega(p,q) \leq 1$, is the weight of the pixel q, defined as a function of the similarity distance between the patch centered on q and Ψ_p : $\omega(i,j) = \frac{1}{Z(i)} \exp\left(-\frac{\|\Psi_i - \Psi_j\|_2^2}{h^2}\right)$. While $Z(i)$ is a normalizing term to validate the constraint $\sum_q \omega(p,q) = 1$. The value of h which controls the decay of the exponential function may have an important impact on the denoising results. So, we first compute, for each similarity metric, the value of h that provides the highest PSNR average on 10 natural images. Figure 4 shows that Bh_2 provides usually the highest PSNR. In terms

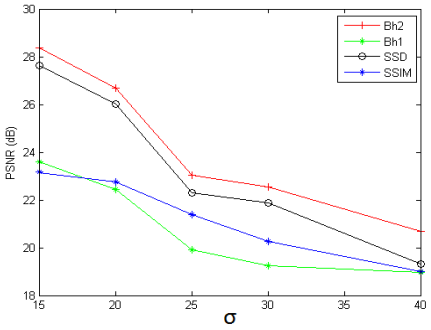


Fig. 4: PSNR values of NLM denoising results using different similarity metrics and the best value of h.

of subjective quality,denoising results of Barbara noisy image (figure 5.b.) show that Bh_2 provides much better noiseless image quality (figure 5.f) than SSD. The PSNR values of denoising results of standard test images summarized in table 1 show that even with a higher noise ($\sigma = 40$) this metric provides an average gain of $1.3dB$ compared to SSIM, SSD and Bh_1. This conclusion is expected since by adding the probability distributions to the SSD, Bh_1 and Bh_2 provide more precise decisions on the similarity between patches that considers both the structure and texture similarities.

5. CONCLUSION

This paper presents the performance of several similarity metrics. A subjective experiment involving 12 observers suggests that SSD which is a widely used similarity metric is not necessarily the most correlated with subjective scores. By using a simple similarity metric combining SSD and the Bhattacharyya coefficient (Bh_2), the performance of the denoising algorithm is improved showing better image quality compared to SSD and SSIM.

Acknowledgment: This work is supported by the ANR-09-VERS-019-02 grant (ARSSO project).

[2]We use the Gabriel Peyré's matlab implementation of NLM algorithm available at http://www.mathworks.com/matlabcentral/fileexchange/13619/

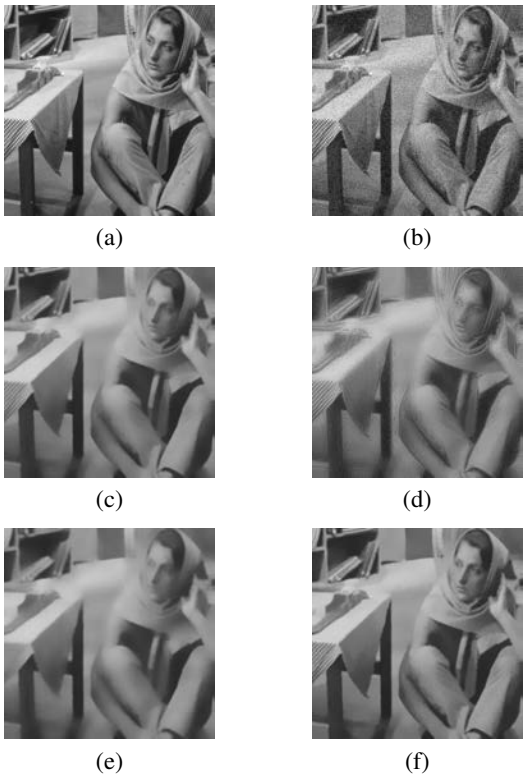


Fig. 5: NLM denoising results using different similarity metrics : (a) original image ; (b) noisy image with $\sigma = 25$; (c) denoised image using SSD; (d) denoised image using SSIM; (e) denoised image using Bh_1; (f) denoised image using Bh_2.

	SSD	SSIM	Bh_1	Bh_2
barbara	19.74	19.37	18.88	20.87
boat	20.01	19.32	19.41	20.93
C.Man	18.47	18.26	17.65	19.5
couple	19.95	19.48	19.45	20.8
F.print	16.82	18.13	15.86	18.05
Lena	20.04	19.36	20.15	21.32
house	20.11	19.48	20.26	22.29
F16	21.08	20.12	21.2	23.39
Peppers	19.03	18.17	19.15	20.42
Babon	17.76	18.21	17.8	19.14
Mean	19.3	18.99	18.98	**20.67**

Table 1: PSNR values of denoising results ($\sigma = 40$).

6. REFERENCES

[1] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," in *IEEE Signal Processing Magazine*, January 2009, vol. 26, pp. 98–117.

[2] I. Mironica, B. Ionescu, and C. Vertan, "The influence of the similarity measure to relevance feedback," in *European Signal Processing Conference (EUSIPCO)*, August 2012, pp. 1573–1577.

[3] Maria Zontak and Michal Irani, "Internal statistics of a single natural image," in *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 977–984.

[4] S. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," in *International Journal of Mathematical Models and Methods in Applied Sciences*, 2007, vol. 1(4), pp. 300–307.

[5] A. Bugeau, M. Bertalmo, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," in *IEEE TRANSACTIONS ON IMAGE PROCESSING*, October 2010, vol. 19, pp. 2634–2645.

[6] O. Le Meur and C. Guillemot, "Super-resolution-based inpainting," in *European Conference on Computer Vision (ECCV)*, October 2012, pp. 554–567.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Dominant orientation templates for real-time detection of texture-less objects," in *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 2257–2264.

[8] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, June 2005, pp. 886–893.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," in *IEEE TRANSACTIONS ON IMAGE PROCESSING*, April 2004, vol. 13, pp. 600–612.

[10] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," in *SIAM Journal on Multiscale Modeling and Simulation*, 2005, vol. 4, pp. 490–530.