

An Efficient and Reliable Image Retrieval Scheme Using STSIM

T.Ammulu¹, G.V.Satya kumar²

¹M.Tech Student of Vasireddy Venkatadri Institute Of Technology, Nambur, Guntur, AP-India,

²Associate Professor, ECE Dept., Vasireddy Venkatadri Institute Of Technology, Nambur, Guntur, AP-India

Email: ¹tammulu471@gmail.com. ²evsk 429@rediffmail.com.

ABSTRACT:

The main objective of this paper is to develop a structural texture similarity metric for image analysis and retrieval. The interest in metrics that deviate from point-by-point similarity was stimulated by the introduction of the structural similarity metrics (SSIM), a class of metrics that attempt to incorporate “structural” information in image comparisons. The proposed metric allow substantial point-by-point deviations between textures that according to human judgment are essentially identical. The proposed metric is computed globally or in sliding windows and implemented using steerable filter decomposition. The performance of the proposed metric is evaluated in the context of “known-item-search”. Experimental results shows that the proposed metric outperform existing metrics PSNR, SSIM. Further the performance of the metric is enhanced by using compression technique with more reliable designs.

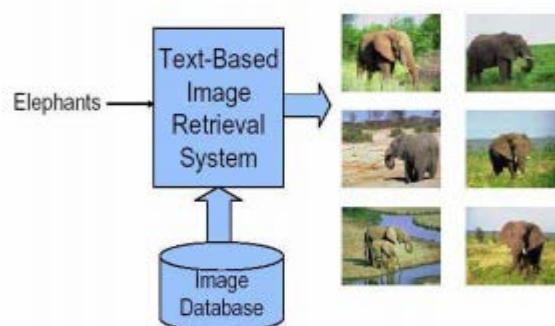
Key Words: Structural Similarity, Image Comparisons, Substantial Point-By-Point Deviations, Sub-Band Statistics, Steerable Filter Decomposition.

I. Introduction

Image retrieval

As human being get image, sound and any other information by seeing, hearing, perception and analysis. Humans can judge similarity of images and sounds according to their semantic contents, instance the searching for a star's picture is based on his facial characters or other contents. So the retrieval methods based on text or keywords for the digital multimedia apparently can't meet the demand that human being get multimedia information exactly [1]. With more and

more multimedia information appear on the Internet and other digital multimedia as



well as human beings thirst for exact and fast retrieval, Image retrieval based on the contents is one of the important aspect of multimedia information retrieval.

Retrieval by using text firstly, in fact it has already changed image retrieval into traditional keywords retrieval is considered as the retrieval based on artificial notes. There are two problems remain in this method. One is the, it brings too heavy workload. On the other hand, it still remains subjectivity and uncertainty. Because of these two drawbacks the image retrieval that is based on artificial notes still remains insufficiency, the further study that adapts vision image features has been coming and become the main study. The character of this method is image feature extraction which is impersonal, whether the retrieval is good or not depends on the accuracy of the features extraction. So the research based on vision features is becoming the focus in the academic community. The feature of vision can be classified by semantic hierarchy into middle level feature and low- level feature. Low-level feature includes color, texture and inflexion. Middle level involves shape description and object feature .

Text-Based approach image is retrieved by the Index images using keywords .The main advantage of this method is easy to implement with fast retrieval. On the other side this method has some drawbacks too. Manual annotation is not always available. Manual annotation is impossible for a large DB and they are not accurate.

Content-Based approach images were retrieved by using the contents of the images .The Advantages Of these approach are Visual features, such as color, texture, and shape information, of images are extracted automatically. The Similarities of images are based on the distances between features. Several CBIR systems currently exist, and are being constantly developed examples are QBIC or Query by Image Content. It was developed by IBM, Almaden Research Centre, to allow users to graphically pose and refine queries based on multiple visual properties such as colour, texture and shape. It supports queries based on input images, user-constructed sketches, and selected colour and texture patterns.

This paper is organized as follows. In chapter-II, gives the review of literature, In chapter-III the proposed improvements and algorithm of this paper are explained results are given in chapter-IV and chapter-V conclusion of this metric.

II. Previous Techniques

Similarity metrics

Similarity metric is a function that measures similarity between two images. This similarity measurement is done by calculating the distance metric(Euclidean,

sum of absolute, etc) between the statistics of the images. If the distance is less then are considered as the similar images. Mean, variance correlation etc can be taken as the statistics.

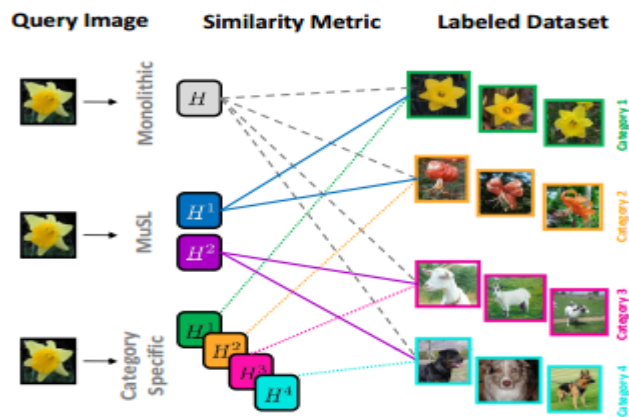
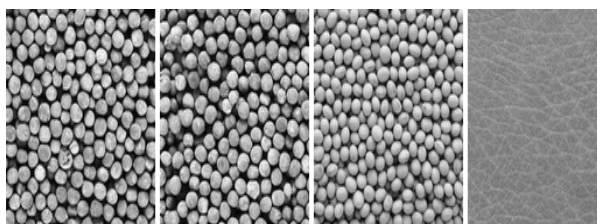


Figure 2:similarity metrics

In the literature several similarity metrics are considered some of them are PSNR, SSIM.

(a) Point-by-Point similarity metrics

In this method the similarity is evaluated by point-by-point basis. This range from simple mean square error(MSE)and peak signal to noise ratio(PSNR).The main disadvantage of this metric is that the value of PSNR decreases with increase in the similarity. This can be shown in Fig.3.



Original PSNR: 8.2dB PSNR: 9.0dB PSNR: 11.3dB

Fig-3: point-by-point metric

(b). Structural similarity metrics

The quality metrics range from MSE and PSNR to more sophisticated metrics that are low level models of human perception. The metrics are typically aimed at near threshold applications such as image compression. That is the lossy compression, there is a need for metric that can be give high similarity scores. This was the first motivation in the development of SSIM. This SSIM metric allows non-structural contrast and changes intensity. The metric SSIM attempt to incorporate "structural" information in image comparison. SSIM metric implemented in either space domain or in wavelet domain. It compare two images or two windows x and y . The SSIM metric can be applied to the images and sub-band images i.e. x , y , x^m and y^m . Here x and y are images and x^m and y^m are sub-band images. First SSIM fixes a window size and shape, with in the images as well as to set the window position. Then it perform the following steps.

The luminance comparison term is defined as

$$l_{x,y}^m = \frac{2\mu_x^m \mu_y^m + c_0}{(\mu_x^m)^2 + (\mu_y^m)^2 + c_0}$$

where μ_x and μ_y are the means of the two windows the contrast comparison term is defined as

$$c_{x,y}^m = \frac{2\sigma_x^m \sigma_y^m + c_1}{(\sigma_x^m)^2 + (\sigma_y^m)^2 + c_1}$$

where σ_x and σ_y are the variances of the two windows and the structure term is defined as

$$s_{x,y}^m = \frac{\sigma_{xy}^m + c_2}{\sigma_x^m \sigma_y^m + c_2}$$

where σ_{xy} is the covariance between the two windows. C_0 , C_1 , and C_2 are small constants.

Finally, this metric combines the three terms into the similarity value

$$q_{ssim}^m(x, y) = (l_{x,y}^m)^\alpha (c_{x,y}^m)^\beta (s_{x,y}^m)^\gamma$$

α , β & γ are the positive weights. Typically set to 1. Sub-band analysis is needed to model early processing in the HVS (human visual system), statistical analysis is necessitated by the stochastic nature of texture similarity, the steerable filter SSIM was implemented. Further it replaces point-by-point comparisons with comparison of region statistics. Thus in order to embrace the SSIM of relying on local image statistics, and to develop a metric that can address the peculiarities of the texture similarity problem. The proposed metric need to completely eliminate point-by point comparisons by

dropping the “structure” term, and to replace it with additional statistics. To overcome some of the limitations of SSIM when applied to texture analysis applications, Zhao et al proposed a STSIM metric.

III Proposed Technique

STRUCTURAL TEXTURE SIMILARITY METRIC (STSIM-1):

This metric was replaced the “structure” term in the CW-SSIM with terms that compare autocorrelations of neighboring subband coefficients in order to provide additional structural and direction.

The first order autocorrelation coefficients can be computed as empirical averages, in horizontal and vertical direction as

$$p_x^m(1,0) = \frac{E\left\{\left[X^m(i,j) - \mu_x^m\right]\left[X^m(i+1,j) - \mu_x^m\right]^*\right\}}{(\sigma_x^m)^2}$$

$$p_x^m(0,1) = \frac{E\left\{\left[X^m(i,j) - \mu_x^m\right]\left[X^m(i,j+1) - \mu_x^m\right]^*\right\}}{(\sigma_x^m)^2}$$

STSIM-1 metric did not use diagonal and anti-diagonal because they did not contribute to any significant improvement in metric performance. The values for autocorrelation are bounded and in the

interval $[-1,1]$, and they are computed for two images as

$$C_{x,y}^m(0,1) = 1 - 0.5 |\rho_x^m(0,1) - \rho_y^m(0,1)|^p$$

$$C_{x,y}^m(1,0) = 1 - 0.5 |\rho_x^m(1,0) - \rho_y^m(1,0)|^p$$

For each sliding window, the previously defined L and C terms are combined with the new statistics into the STSIM as

$$q_{\text{STSIM-1}}^m(x, y) = (L_{x,y}^m)^{\frac{1}{4}} (C_{x,y}^m)^{\frac{1}{4}} (C_{x,y}^m(0,1))^{\frac{1}{4}} (C_{x,y}^m(1,0))^{\frac{1}{4}}$$

The STSTM-1 shows better performance i.e. closer to human judgment of texture similarity than SSIM and CW-SSIM.

Selection of sub-bands:

In this section metrics were developed that extends the ideas including a boarder set of images statistics. The work that was proposed by portilla and simoncelli[16] states that a broad class of textures can be synthesized using a set of statistics and characterize the coefficients of a multiscale frequency decomposition.

This method adopts the mean and variance of the original SSIM metric in addition with the crossband correlation between the subbands. The reason for adding crossband correlations is that the image representation by the steerable filter is over complete, so the subbands

coefficients are correlated. The crossband correlation statistics are computed based on magnitudes of the coefficients. The raw complex coefficients may in fact be uncorrelated, since phase information can lead to cancellations. Simoncelli shows that, the magnitudes of the wavelet coefficients are not statistically independent and large magnitudes in subbands of natural images tend to occur at the same spatial locations in subbands at adjacent scales and orientations. The intuitive explanation may be that the “visual” features of natural images do give rise to large local neighborhood spatial correlations, as well as large scale and orientation correlations[16].

The crossband-correlation coefficient between subbands m and n (excluding the lowpass and highpass bands) is computed as:

$$p_{|x|}^{m,n}(0,0) = \frac{E\{(|x^m(i,j) - \mu_{|x|}^m|)(|x^n(i,j) - \mu_{|x|}^n|)\}}{\sigma_{|x|}^m \sigma_{|x|}^n}$$

This will be essential when two images were compared for compression and image quality, here we ignore the point-by-point comparisons but the local variations are penalized by the metrics. Window size determines the texture scale. Note that the window size determines the texture scale that is relevant to our

problem. Thus if the window is large enough to include several repetitions of the basic texture pattern, e.g., several peas, then the peace of treated as a texture, Otherwise the metric is will focus on the surface texture of the individual peas.

Steerable Filter Decomposition:

The steerable filter to decompose the real image x and complex sub-band x^m . The real and imaginary parts of those sub-bands are not independent to each other, the imaginary part is the Hilbert transform of the real part. The quadrature filters are used to envelope the detection and feature extraction in images. In this metric 3-scale 4-orientation steerable filter is used to decompose the image. Complex or real steerable filter is applied to the global window. When the real or complex steerable filter are applied on the global window the performance is same, where as it differs in the local window.

STRUCTURAL TEXTURE SIMILARITY METRIC(STSIM-2):

The metric will use the mean value $|\mu_x^m|$ variance $(\sigma_x^m)^2$, and autocorrelations $p_x^{m,n}(0,1)$ and $p_x^{m,n}(1,0)$ computed on the complex subband coefficients, and the cross-band correlation $p_{|x|}^{m,n}(0,0)$ defined on

the magnitudes. If we adopt the SSIM approach for comparing image statistics, all we need to do is add a term for comparing the crossband-correlation coefficients to the STSIM-1 metric. Like the STSIM-1 comparison terms, this term should take into account the range of the statistic values and should also produce a number in the interval $[0, 1]$:

$$c_{x,y}^{m,n}(0,0) = 1 - 0.5 \left| p_{|x|}^{m,n}(0,0) - p_{|y|}^{m,n}(0,0) \right|^p$$

Again, typically, $p = 1$.

Note that since the crossband correlation comparison terms involve two subbands, it does not make sense to multiply them with the other STSIM-1 terms in (12). We thus need a separate term. For a given window, the overall STSIM-2 metric can then be obtained as a sum of two terms: one that combines the STSIM-1 values over all subbands, and one that combines all the crossband correlations.

$$q_{STSIM\ 2}(x, y) = \frac{\sum_{m=1}^{N_b} q_{STSIM\ 1}^m(x, y) + \sum_{i=1}^{N_c} c_{x,y}^{m_i, n_i}(0,0)}{N_b + N_c}$$

When the metric is applied on a sliding window as is, spatial pooling is needed to obtain an overall metric value QSTSIM-2(x, y). As we saw above, spatial pooling

can be done before or after the summation. finally this project is enhanced by using compression technique with more reliable designs.

Algorithm:

- 1) A multiscale frequency decomposition: Such decompositions can be real or complex. In the following, we will use the three-scale, four-orientation steerable filter decomposition as in CW-SSIM.
- 2) A number of subband statistics: Each statistic corresponds to one image and is computed within one window in that image. Statistics are computed within a subband or across subbands.
- 3) The window over which the statistics are computed can be local (sliding window) or global (the entire image).
- 4) A means for comparing (corresponding) subband statistics, one from each image whose similarity we wish to assess: The particular formula depends on the range of values that the statistic takes, and yields a nonnegative number that represents the similarity or dissimilarity of the two statistics.

IV Results

This is the comparison table of existing technique

Images	Existing Time	Enhanced Time with decomposition level:2
Tiger	22.840882 seconds	17.525739 seconds.
Russian Girl	20.249149 seconds	17.441487 seconds.
Flower Pot	20.362910 seconds.	17.367584 seconds.

When an image is given as a query image, this metric checks for the similar images in the database

QUERY IMAGE



Fig-4:query image

Original image



Compressed image



Reconstructed image



Fig -5:output of the compressed and reconstructed images

As the large images are not recognized the query image is compressed and reconstructed. To retrieve the image the decomposition level is given.

Compression Ratio:51.3468

DATABASE IMAGE



Fig-6:Retrieved image

V Conclusion

In this paper, a new metric was developed to determine structural texture similarity. The proposed metric accounts the human visual perception and stochastic nature of texture. In this method steerable filter is used to decompose the image into sub-bands. Compared to the existing techniques the proposed metric uses the cross-band correlation of sub-bands to improve the overall efficiency. This metric enables the comparison with the human performance on a large database which eliminates the need for cumbersome subjective tests. By inserting the compression module more sophisticated design was implemented. The proposed

metric gives better results compared to the existing metrics like PSNR, SSIM.

VI. REFERENCES

- [1] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, pp. 177–200, 1998.
- [2] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Academic Press, 2005, pp. 939–959.
- [3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [4] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. II, Philadelphia, PA, 2005, pp. 573–576.
- [5] X. Zhao, M. G. Reyes, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for retrieval applications," in *Proc. Int. Conf. Image*

Processing (ICIP), San Diego, CA, Oct. 2008, pp. 1196–1199.

[6] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, “Structural similarity metrics for texture analysis and retrieval,” in Proc. Int. Conf. Image Processing, Cairo, Egypt, Nov. 2009, pp. 2225–2228.

[7] D. Cano and T. H. Minh, “Texture synthesis using hierarchical linear transforms,” *Signal Processing*, vol. 15, pp. 131–148, 1988.

[8] M. Porat and Y. Y. Zeevi, “Localized texture processing in vision: Analysis and synthesis in Gaborian space,” *IEEE Trans. Biomed. Eng.*, vol. 36, no. 1, pp. 115–129, 1989.

[9] K. Popat and R. W. Picard, “Novel cluster-based probability model for texture synthesis, classification, and compression,” in Proc. SPIE Visual Communications '93, Cambridge, MA, 1993.

[10] D. J. Heeger and J. R. Bergen, “Pyramid-based texture analysis/synthesis,” in Proc. Int. Conf. Image Processing (ICIP), vol. III, Washington, DC, Oct. 1995, pp. 648–651.

[11] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint

statistics of complex wavelet coefficients,” *Int. J. Computer Vision*, vol. 40, no. 1, pp. 49–71, Oct. 2000.

[12] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multi-scale transforms,” *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[13] G. Jin, Y. Zhai, T. N. Pappas, and D. L. Neuhoff, “Matched-texture coding for structurally lossless compression,” in Proc. Int. Conf. Image Processing (ICIP), Orlando, FL, Oct. 2012, accepted.

[14] C. T. Meadow, B. R. Boyce, D. H. Kraft, and C. Barry, *Text information retrieval systems*. Emerald Group Publishing, 2007.

[15] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance,” *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[16] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int. J. Computer Vision*, vol. 40, no. 1, pp. 49–71, Oct. 2000.