

Virus Texture Analysis Using Local Binary Patterns and Radial Density Profiles

Gustaf Kylberg¹, Mats Uppström², and Ida-Maria Sintorn¹

¹ Centre for Image Analysis, Lägerhyddsvägen 2, SE-751 05 Uppsala, Sweden
{gustaf.kylberg, ida.sintorn}@cb.uu.se

² Vironova AB, Gävlegatan 22, SE-113 30 Stockholm, Sweden
mats.uppstrom@vironova.com

Abstract. We investigate the discriminant power of two local and two global texture measures on virus images. The viruses are imaged using negative stain transmission electron microscopy. Local binary patterns and a multi scale extension are compared to radial density profiles in the spatial domain and in the Fourier domain. To assess the discriminant potential of the texture measures a Random Forest classifier is used. Our analysis shows that the multi scale extension performs better than the standard local binary patterns and that radial density profiles in comparison is a rather poor virus texture discriminating measure. Furthermore, we show that the multi scale extension and the profiles in Fourier domain are both good texture measures and that they complement each other well, that is, they seem to detect different texture properties. Combining the two, hence, improves the discrimination between virus textures.

Keywords: virus morphology, texture analysis, local binary patterns, radial density profiles.

1 Introduction

To image viruses using negative stain transmission electron microscopy (TEM) has proven to be an invaluable tool in early virus diagnostics, [1,2]. Viruses show different surface texture when imaged using TEM. This fact has been utilized from the very beginning of virology when the advances in TEM technology walked hand in hand with the discovery of new viruses and the creation of a virus taxonomy.

The analysis of a virus sample using TEM typically means an visual inspection performed at the microscope. The main problems with this procedure are the need for an expert to perform the analysis at the microscope and that the result is highly dependent on the expert's skill and experience. To make virus diagnostic using TEM more useful, automatic analysis would hence be desirable. The analysis presented in this paper is part of a project with the aim to develop a fully automatic system for virus diagnostics based on TEM in combination with automatic image analysis.

Viruses vary in shape from icosahedral to highly pleomorphic particles and different virus types have different sizes. The appearance of virus particles in

TEM images can be divided into properties like size, shape and texture. While size and shape can be used to exclude some virus types they can not, by themselves, confirm a specific virus type. Many viruses show a distinct and recurring texture making it an interesting property to analyse and use to discriminate between different virus types.

Very little work has been reported on analysing virus texture in TEM images. In [3] ring filters in the Fourier power spectrum are used as features to discriminate between four icosahedral viruses. In [4] higher order spectral features are utilized to differentiate between the same four icosahedral viruses. There is no consensus definition of what texture is, but the general opinion often include some repetitive intensity variation. The definition of a good texture measure is hence highly dependent on the problem at hand. When a measure is used in a classification procedure it becomes possible to assess its capabilities and qualities.

Local binary patterns (LBP) emerged in the mid '90s as a local texture measure [5,6]. LBP has along with several extensions become a popular texture measure in several real-world applications, see e.g., [7,8]. The thesis by Mäenpää [9] gives a good overview of LBP and some of its extensions.

Another way of describing the intensity variations in an object is to compute a radial or density profile (RDP). In [10] radial density profiles are used to discriminate between three maturation stages of human cytomegalovirus capsids in TEM images of cell sections. [11,12] are examples of their use in cryo-electron microscopy where DNA packing is compared between two virus types and attachment sites on Simian Cytomegalovirus capsids are analysed, respectively.

In this paper the basic concepts of LBP and RDP, along with some variations, are investigated for the problem of discriminating between virus textures. We use Random Forest [13], an ensemble classifier based on decision trees, to enable a quantitative comparison of how the different texture measures can discriminate between 15 virus types.

2 Material

The data set consists of 15 different virus types represented with 100 TEM image patches each. The virus types are of different sizes and shape. However, the diameter (most common cross section for non spherical viruses) is relatively constant within a virus type. The virus types range from 25 to 270 nm in diameter and their shapes vary from icosahedral to highly pleomorphic (for example like boiled spaghetti). The image patches are disk shaped cutouts centred on automatically segmented virus particles using the segmentation method presented in [14]. The viruses have been imaged at different magnifications in the TEM with a pixel size ranging from 0.5 to 5 nm. To get comparable texture samples we resample the images to two specific scales using bilinear interpolation. In the first, which we call *fixed scale*, the size of a pixel is 1 nanometer. In the second, which we call *object scale*, the radius of a virus particle is represented by 20 pixels. In Fig. 1 an image patch of each virus type in the object scale is shown along with the virus name and diameter.

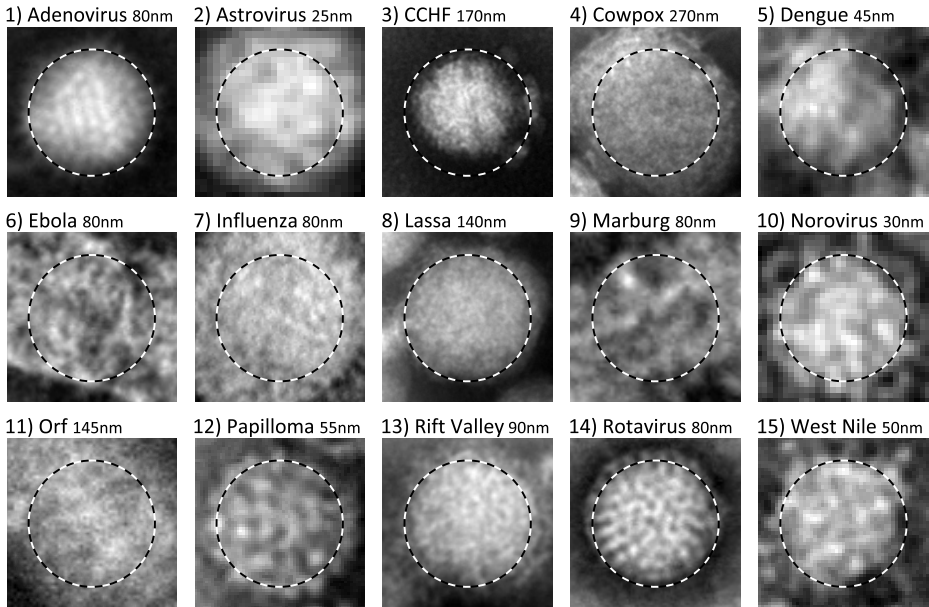


Fig. 1. Example images of the 15 virus types in the data set resampled to object scale. Following the virus name is the approximate particle diameter of each virus type. The dashed circle marks the texture patch used in the analysis.

3 Methods

3.1 LBP

The local binary pattern (LBP) at a pixel q_c with the position (x_c, y_c) in an image I is computed by thresholding a number, N , of neighbour points, p , evenly distributed at a radius R around q_c . The position of the neighbour point p where $p \in [0, \dots, N - 1]$, is given by: $(x_c + R \cos(2\pi p/N), y_c - R \sin(2\pi p/N))$.

The LBP code is then the sequence of zeros and ones from the thresholded values in the neighbour points. If a point p does not coincide with a pixel centre, bilinear interpolation is used to compute the gray value in p . The LBP code can be made rotational invariant, LBP^{ri} , by circularly shifting the binary number until the minimum value is obtained. Furthermore we can restrict our binary codes, considering only uniform binary patterns, LBP^{riu} , further limiting the number of possible codes. Uniform binary patterns are patterns with at the most two spatial transitions between 0 and 1 or 1 and 0. Detailed definitions of LBP, rotational invariance and uniformity can be found in [9]. The rotational invariant and uniform (allowing ≤ 2 transitions) LBP using N samples at the radius R is denoted $\text{LBP}_{N,R}^{\text{riu}2}$. The LBP measure of a set of pixels is then the histogram of occurring LBP codes.

3.2 LBPF

Extending LBP to multiple scales opens up many design options, whereof several alternatives are presented in [9]. A straight forward extension of LBP is to compute several $\text{LBP}_{N,R}^i$ with increasing R where the points are sampled using Gaussian kernels. This is denoted $\text{LBPF}_{N,R}^i$, where the additional F stands for filtered. The standard deviation of the Gaussian kernels and the sample points are selected to cover the neighbourhood as well as possible while minimizing the overlap of kernels. We have used an exponentially growing radius R and Gaussian kernels computed as described in [9]. The LBPF codes are then concatenated into one feature vector.

3.3 RDP

The radial mean intensity, f , at radius r from the center pixel q_c in an image I is defined as:

$$f(q_c, r) = \frac{1}{|N|} \sum_{q \in N} I(q), \quad (1)$$

$$N = \{q : \|q - q_c\|_2 \in (r - 0.5, r + 0.5]\}, \quad (2)$$

where q is a pixel at radius r from q_c and N is the set of pixels at radius r from q_c . $|N|$ is the number of pixels in the set N . The radial density profile with n radii, RDP_n , computed for the pixel q_c is:

$$\text{RDP}_n = [f(q_c, 1) - \bar{f}_{q_c} \quad f(q_c, 2) - \bar{f}_{q_c} \quad \dots \quad f(q_c, n) - \bar{f}_{q_c}], \quad (3)$$

where \bar{f}_{q_c} is the mean value of all $f(q_c, r)$, $r \in [1, 2, \dots, n]$.

3.4 FRDP

The FRDP is computed in the same way as the RDP but using the Fourier magnitude spectra in a \log_e scale as the input image. In this way, the FRDP shows a profile of frequencies occurring in the input image I . FRDP can be interpreted as a generalization of the spectral rings used in [3].

3.5 Classification

To get objective measures of the performances of the investigated texture measures we use the Random Forest classifier. That is an example of an ensemble classifier based on bagged decision trees introduced by Breiman in [13]. When the ensemble is created a new bootstrap sample is drawn for each new tree. When a tree is grown only a random subset of the feature values are used, increasing the diversity among trees even further. The error rate of the built ensemble classifier can be estimated through the samples left out of the bootstrap samples, called “out-of-bag” data by Breiman. We grew 200 trees and the increase in performance per added tree levelled out between 100 and 200 trees meaning that 100 trees is a large enough ensemble. The number of feature values to select at random for each decision split is set to the square root of the number of feature values, proposed by Breiman as a rule of thumb.

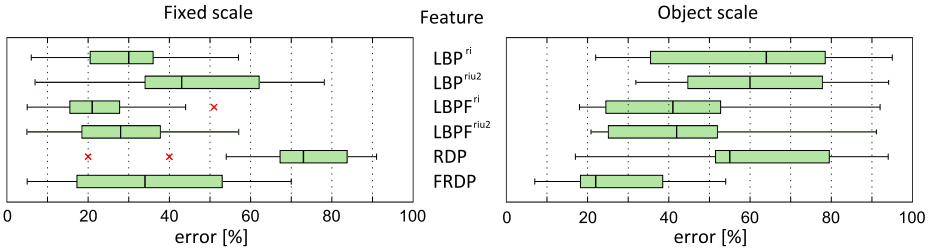


Fig. 2. Estimated classification errors for the Random Forest classifier for the different texture measures. The boxes stretch from the lower to the upper quartile and the line marks the median. The whiskers show min and max in the data excluding outliers. Outliers (x) are data points at least 1.5 times the size of the box away from the box.

4 Results

For the investigation in this paper we have compared i) $LBP_{8,2}^{ri}$, ii) $LBP_{8,2}^{riu2}$, iii) a multi scale LBP composed of $LBP_{8,1}^{ri} + LBP_{8,2,4}^{ri} + LBP_{8,5,4}^{ri}$, iv) the uniform variant $LBP_{8,1}^{riu2} + LBP_{8,2,4}^{riu2} + LBP_{8,5,4}^{riu2}$, v) RDP_{20} and vi) $FRDP_{20}$. The latter two measures are of global character while the LBP variations are of local character. Figure 2 shows the result. We found that the LBP^{ri} in fixed scale and the $FRDP$ in object scale are the two most promising texture measures investigated.

For $LBP_{N,R}$ a range of parameter values were tested: $N \in [4, 8, 16]$, $R \in [1, 2, 3, 4]$. For LBP the following sets of N were tested: $N_1 = [4, 8, 8]$, $N_2 = [4, 8, 16]$, $N_3 = [8, 8, 8]$, $N_4 = [8, 8, 16]$ were tested together with R calculated according to [9]. Many parameter combination resulted in similar discriminant power and the values selected were the best performing options.

The result shows that LBP and its variations generally performed better in the fixed scale. From Fig. 2 it is clear that applying the uniformity restriction results in a poorer discrimination between the virus types in the fixed scale. Among the 1,500 texture patches in the data set, $LBP_{8,2}^{riu2}$ resulted in 263 more samples being wrongly classified in the fixed scale compared to using $LBP_{8,2}^{ri}$. Applying the uniformity restriction in the LBP also resulted in a poorer classification result in the fixed scale, but only with 93 more samples incorrectly classified. Most prominent, removing the uniformity restriction for samples of the Cowpox virus the error decreased from 54% to 26% showing that for certain virus textures important discriminant information is found in the non uniform patterns.

Figure 3 shows the confusion matrices for the classification using LBP, LBP, RDP and FRDP for the fixed scale (a), and the object scale (b). These matrices display the information constituting the boxes for these texture measures in Fig. 2. From the figures it is clear that RDP discriminates between the virus textures rather poorly (weak diagonal and relatively high values everywhere in the matrix). It is also easy to see that both LBP and LBP perform better in the fixed scale (generally lower off-diagonal values), and that the opposite is true for RDP. Both Fig. 3 and Fig. 2 show that the texture measures best discriminating the virus

textures are LBP^{ri} in the fixed scale and RDPF in the object scale. By carefully analysing their confusion matrices one can see that LBP^{ri} clearly perform better for certain virus textures e.g. 4, 6 and 12, whereas RDPF perform better for 15. Combining these two measures would probably give an even better discrimination. That is in fact the case which is shown in Fig. 4. In the confusion matrix, Fig. 4. a), the off-diagonal values are lower, and the values on the diagonal are much higher compared to the confusion matrices for each of the two texture measures. It is also

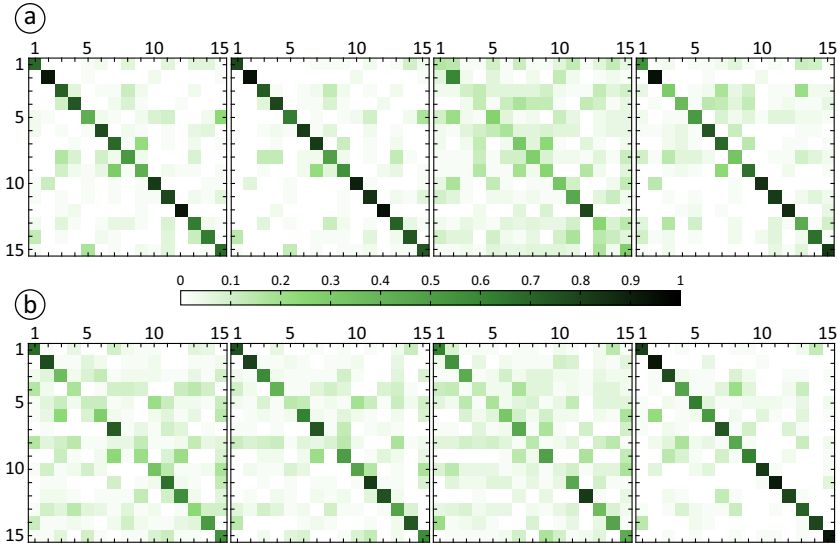


Fig. 3. Resulting confusion matrices, i.e., from the Random Forest classification for the texture measures, from left to right: LBP^{ri}, LBPF^{ri}, RDP, FRDP for a) image patches in fixed scale and b) object scale

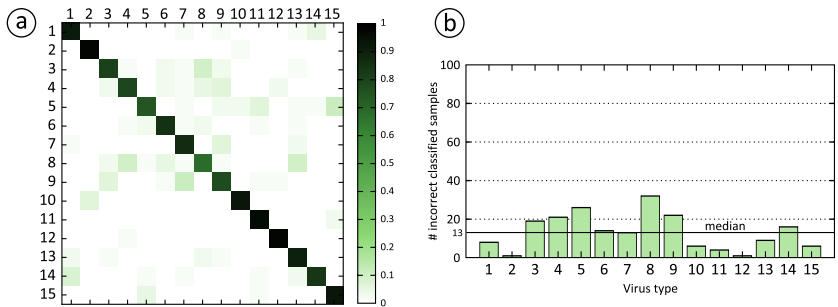


Fig. 4. a) Resulting confusion matrix after combining the LBPF^{ri} measure in the fixed scale with the FRDP in the object scale. b) Total error in classification for each virus type. The median value is marked with a horizontal line.

clear that the two measures provide complementary information for many of the virus textures as using both in the classification results in better discrimination than using the best of the two for each virus class, see e.g. virus 5, 8, 9, and 13. Figure 4 b) shows the classification result from using the combination of the two measures for each of the virus classes. The median classification error rate is 13 % which should be compared to 21% for LBP in the fixed scale and 22% for RDP in the object scale.

5 Discussion

When the uniformity restriction was introduced in [15] the authors show that the discriminant power was mainly made up by the uniform patterns. This restriction is commonly regarded as an improvement upon the basic LBP when applied. However, our results show that the non uniform patterns contribute to the discriminant power of LBP and LBP for the virus texture data.

The approach to handle the different sizes of the viruses was to resample the images into a fixed scale and into an object scale. From Fig. 2 we can conclude that the global measures generally score better in the object scale while the opposite can be observed for the two LBP-based measures.

The choice of classifier can of course be discussed and with a different classifier, e.g., SVM, NN or AdaBoost, the result would most likely have looked slightly different. We have selected the Random Forest classifier based on previous positive experiences using similar measures and on a comparison of classifiers (Random Forest, SVM, GMM, AdaBoost) on a similar problem. However, this paper is not about selecting the best suited classifier but rather using a classifier as a tool to evaluate our texture measures.

Future work includes exploring some of the many possibilities within the LBP framework to make a local texture descriptor that is more robust to noise. Further, combinations of texture measures and size and shape descriptors as well as other *a priori* knowledge about the virus sample type will be used in the discrimination problem present in our intended real-world application. For example haemorrhagic fever viruses such as CCHF, Dengue and Ebola are rarely found in fecal samples. Moreover, many virus particles will be analysed and classified in each patient sample and therefore a certain degree of incorrect classifications do not pose a problem.

Acknowledgement. Thanks to: Ali Mirazimi and Kjell-Olof Hedlund at SMI for preparing the virus samples; Tobias Bergroth and Lars Haag at Vironova for acquiring the images; Erik Wernersson and Gunilla Borgefors for valuable comments regarding the manuscript. The work presented in this paper is part of a project funded by the Swedish Agency for Innovation Systems (VINNOVA), Swedish Defence Materiel Administration (FMV), and the Swedish Civil Contingencies Agency (MSB).

References

1. Goldsmith, C.S., Miller, S.E.: Modern uses of electron microscopy for detection of viruses. *Clin. Microbiol. Rev.* 22(4), 552–563 (2009)
2. Biel, S.S., Madeley, D.: Diagnostic virology – the need for electron microscopy: a discussion paper. *J. Clin. Virol.* 22(1), 1–9 (2001)
3. Matuszewski, B.J., Shark, L.K.: Hierarchical iterative bayesian approach to automatic recognition of biological viruses in electron microscope images. In: *Proc. of 2001 International Conference on Image Processing (ICIP)*, vol. 2, pp. 347–350 (2001)
4. Ong, H.C.L.: Virus recognition in electron microscope images using higher order spectral features. PhD thesis, Queensland University of Technology (2006)
5. Harwood, D., Ojala, T., Pietikäinen, M., Kelman, S., Davis, L.: Texture classification by center-symmetric auto-correlation, using kullback discrimination of distributions. *Pattern. Recogn. Lett.* 16(1), 1–10 (1995)
6. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern. Recogn.* 29(1), 51–59 (1996)
7. Hervé, N., Servais, A., Thervet, E., Olivo-Marin, J.C., Meas-Yedid, V.: Statistical color texture descriptors for histological images analysis. In: *Proc. of IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 724–727 (2011)
8. Zhang, B.: Classification of subcellular phenotype images by decision templates for classifier ensemble. In: Pham, T., Zhou, X. (eds.) *Proc. of 2009 International Conference on Computational Models for Life Sciences (CMLS)*, pp. 13–22 (2010)
9. Mäenpää, T.: The local binary pattern approach to texture analysis - extensions and applications. PhD thesis, University of Oulu (2003)
10. Sintorn, I.M., Homman-Loudiyi, M., Söderberg-Nauclér, C., Borgefors, G.: A refined circular template matching method for classification of human cytomegalovirus capsids in TEM images. *Comput. Meth. Prog. Bio.* 76, 95–102 (2004)
11. Bhella, D., Rixon, F.J., Dargan, D.J.: Cryomicroscopy of human cytomegalovirus virions reveals more densely packed genomic DNA than in herpes simplex virus type 1. *J. Mol. Biol.* 295, 155–161 (2000)
12. Trus, B.S., Gibson, W., Cheng, N., Steven, A.C.: Capsid structure of Simian cytomegalovirus from cryoelectron microscopy: Evidence for tegument attachment sites. *J. Virol.* 73(3), 2181–2192 (1999)
13. Breiman, L.: Random forests. *Machine Learning* 45, 5–32 (2001)
14. Kylberg, G., Uppström, M., Hedlund, K.O., Borgefors, G., Sintorn, I.M.: Segmentation of virus particle candidates in transmission electron microscopy images (manuscript, 2011)
15. Mäenpää, T., Ojala, T., Pietikäinen, M., Soriano, M.: Robust texture classification by subsets of local binary patterns. In: *Proc. of International Conference on Pattern Recognition (ICPR)*, pp. 3947–3950 (2000)