GENERALIZED EXTREME VALUE DISTRIBUTIONS, INFORMATION GEOMETRY AND SHARPNESS FUNCTIONS FOR MICROSCOPY IMAGES

Reiner Lenz

Linköping University
Department Science and Technology
Bredgatan SE-60174 Norrköping, Sweden

ABSTRACT

We introduce the generalized extreme value distributions as descriptors of edge-related visual appearance properties. Theoretically these distributions are characterized by their limiting and stability properties which gives them a role similar to that of the normal distributions. Empirically we will show that these distributions provide a good fit for images from a large database of microscopy images with two visually very different types of images. The generalized extreme value distributions are transformed exponential distributions for which analytical expressions for the Fisher matrix are available. We will show how the determinant of the Fisher matrix and the gradient of the determinant of the Fisher matrix can be used as sharpness functions and a combination of the determinant and the gradient information can be used to improve the quality of the focus estimation.

Index Terms— generalized extreme value distribution, information geometry, edge statistics, auto-focus, image-based screening

1. OVERVIEW AND BACKGROUND

Edge detection is one of the first and most important operations in all technical and biological vision systems. It is therefore important to understand the relation between the statistical properties of the space of input images and the statistical properties of the resulting edge detector values. In this paper we mainly consider images from an automated microscope taking focus series of cells with two different types of staining. We have thus two sets of images with two visually different properties, the optical properties of the system in the form of the focus plane are systematically varied and the optimal focus setting is established with the help of an independent measuring process. Statistical properties of edge detectors have been studied earlier, mainly in the context of natural image statistics (see [1, 2, 3, 4, 5, 6, 7] for some examples). The main observation is that the distributions of these filter results usually have heavy tails, with the (two-parameter) Weibull distribution as the most popular choice. In this paper we will use the generalized extreme value (GEV) distributions which come in three different types: the Fréchet-, (reversed) Weibull- and Gumbel-distributions. We will compare the fitting properties of the GEV distributions with those of the standard Weibull distribution and show that the fitting results for the GEV are comparable or slightly better than those for the Weibull distributions.

The second topic we investigate is the usage of concepts from information geometry [8, 9] where probability distributions are points on a manifold and methods from differential geometry are used to study the relation between different distributions. Typical concepts used are the distance between distributions or the shortest path (the geodesic) between them. Both, Weibull- and GEV-distributions, can be derived by transformations from the exponential distribution and closed form expressions for the metric of these manifolds are therefore available (see [10, 11]). In the general case we can, however, not apply the tools of information geometry directly since the domains on which these distributions are defined vary depending on the value of their parameters. We thus use only local properties following a similar approach used to generalize transformation groups to local transformation groups, see [12]. Using tools from differential geometry we introduce a local metric in the space of the GEV-distributions with the help of the Fisher matrix. For the focus sequences we show that the determinant of the Fisher matrix can be used as a sharpness function. For dynamic focus sequences the change of the probability distributions between consecutive frames can be measured in the Fisher geometry and combining the static sharpness function with the characterization of the dynamic change improves the autofocus results. This approach is similar to the framework developed in [13] where it is shown how geometry-based optimization methods can be used to improve the efficiency of auto-focus control.

The study of efficient implementations or the comparison to existing auto-focus methods is outside the scope of this study. The fact that these images can be described by

This work was supported by the Swedish Foundation for Strategic Research through grant IIS11-0081

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

a parametric model for which analytic expressions of the relevant geometric variables are available is however an advantage compared to methods that completely depend on empirical values such as the data variance.

2. IMAGES, FILTERS AND DISTRIBUTIONS

In the following we will use the microscopy images from set BBBC006v1 in the Broad Bioimage Benchmark Collection which is described in [14]. The images are available at `http://www.broadinstitute.org/bbbc/BBBC006`. The database contains 52224 images from 384 cells, measured at two positions and prepared with two different types of staining. For each position and each cell a focus sequence consisting of 34 images was recorded. This resulted in 384x2x2x34 = 52224 images. The images show stained human cells and the imaging process is desribed as follows: *For each site, the optimal focus was found using laser autofocusing to find the well bottom. The automated microscope was then programmed to collect a z-stack of 32 image sets.* For one cell we show four images in Figure 1. We see focus position 7 in the top and position 16 in the bottom row. Stainings W1 are in the left and W2 in the right column.

In the experiments we use filter kernels defined on a window consisting of 5x5 pixels. The filters are constructed with the help of the representation theory of the dihedral group and correspond to the 2-dimensional irreducible representations of the dihedral group (the details of this construction can be found in [15] and [16]). In the current application the following properties are essential: (1) each filter kernel consist of an equal number of +1 and -1 filter coefficients, (2) the filters come in pairs f_x, f_y, roughly corresponding to an x- and a y-gradient filter, (3) for a 5x5 window there are six such filter pairs and (4) for the filter results e_x, e_y, obtained by applying the two filters f_x, f_y, the squared magnitude $e_x^2 + e_y^2$ is independent under the application of all operations of the dihedral group elements to the underlying window. From the six filter pairs used we obtain six filter magnitude values $m_1, \ldots m_6$ in each pixel and as the final filter result we use their sum $m = m_1 + \ldots + m_6$. From the definition follows that this filter result is the combination of six different edge-like filter responses. We will only use this setup for the filter process so that all statistical evaluations are based on the same pre-processing results. We use these filters since they have a strong theoretical foundation in group representation theory and since they are very fast, consisting of addition and subtraction of pixels values only. For other filter systems (like the Gabor-based Gist filters [17]) we found similar properties.

The original microscope images were scaled-down and then pre-processed with a simple combination of gray-value thresholding and morphological operations to construct a mask in which dark background regions are suppressed. 360 images out of the 52224 images in the database contained too few object points and where therefore excluded from the analysis. In a second series of experiments we constructed the same mask as described above but we replace the original (12-bit) pixel values g by their log transform $g_l = log(1 + g)$.

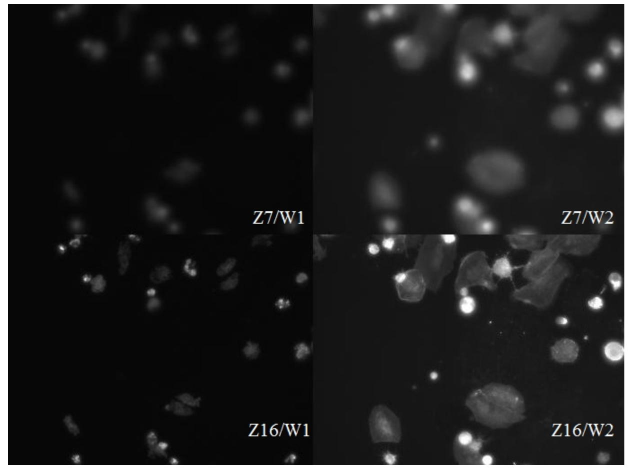


Fig. 1: Images z7/W1&W2,z16/W1&W2, see main text

The magnitude filter results from the pixels under the mask are characterized by GEV-distributions. Theoretically the GEV-or Max-Stable distributions are attractive since they share two main characteristics with the normal distributions: they are limit-distributions and they are stable. The "Extremal Type Theorem" (Theorem 1.4.2 in [18]) states that if a suitably normalized sequence of random variables M_n of the form $M_n = max(\xi_1, \ldots, \xi_n)$ converges to a non-degenerate distribution then that limit-distribution must be of one of the three types of GEV-distributions mentioned below. This is a result that corresponds to the central limit theorem that state that the sum of random variables converges to the normal distribution. The second characteristic of the normal distributions is that the sum of two normally distributed random variables is also normally distributed. If we replace the sum with the maximum operation then we obtain the max-stable distributions and a fundamental theorem (Theorem 1.4.1 in [18]) states that max-stable distributions are GEV-distributions and vice versa. Practically we can expect that the additional third parameter adds more flexibility to two-parameter models like the two-parameter Weibull distributions and that we can therefore expect better fitting results. Note that pixels that contribute to the distribution fitting are selected using a masking procedure. We thus have an indirect thresholding process suppressing positions with low filter results.

The GEV-distributions are characterized by three parameters: location (μ), scale (σ) and shape (ξ) and the three parameter Weibull distribution is given by the (different) parameters location (μ), scale (β) and shape (α). Depending on the value of the shape parameter ξ the GEV-distributions are known as: Gumbel distributions ($\xi = 0$), Fréchet distributions ($\xi > 0$) and Weibull distributions ($\xi < 0$). More information about these distributions (and the sometimes confusing naming conventions) can be found in [18]. It is often argued that GEV-distributions are more flexible since they "let the data decide" which of the models fits best. In the following the distribution parameters are estimated from measured data with the help of maximum-likelihood estimators.

We write $p(x;\theta) = p_\theta(x)$ to denote the probability density function (pdf) of a distribution parametrized by the parameter vector θ and $F(x;\theta) = F_\theta(x)$ for the corresponding cumulative density function (cdf). For a stochastic variable $h(x)$ and a distribution $p_\theta(x)$ we write the expectation of h as: $\mathcal{E}_\theta(h)$. In this notation the cdfs of the GEV distributions (with positive shape parameter ξ) are given by:

$$F(x;\mu,\sigma,\xi) = \begin{cases} \mathrm{e}^{-\left(\frac{\xi(x-\mu)}{\sigma}+1\right)^{-1/\xi}} & \text{if } \frac{\xi(x-\mu)}{\sigma}+1>0; \\ 0 & \text{else.} \end{cases}$$

The cdf for the standard three-parameter Weibull distribution is $F(x;\mu,\beta,\alpha) = 1 - \mathrm{e}^{-\beta^{-\alpha}(x-\mu)^\alpha}$ for $x > \mu$. It follows that the support of these distributions varies depending on the parameter values and the standard assumptions of information geometry are therefore not applicable.

For parametric probability distributions with general parameter vector $\theta = (\theta_1, ..., \theta_K)$ and pdf $p(x;\theta)$ the Fisher information matrix $\mathcal{G}(\theta) = (g_{kl})$ with elements g_{kl} is defined as

$$g_{kl} = \int_x \frac{\partial \log p(x;\theta)}{\partial \theta_k} \frac{\partial \log p(x;\theta)}{\partial \theta_l} p(x;\theta)\, dx \quad (1)$$

The properties of the Fisher matrix make it theoretically attractive but it is often difficult to compute analytically. For the Weibull and the GEV-distributions the situation is different since they can be derived by transformations from the exponential distribution (see [10, 11, 19]). We note that μ is only a shift parameter and the Fisher matrix is therefore a function of the variables σ and ξ only. Symbolic expressions of the partial derivatives of the determinant of the Fisher matrix were computed with the help of Mathematica. We use only local properties following a similar approach used to generalize transformation groups to local transformation groups, see [12].

3. EXPERIMENTS

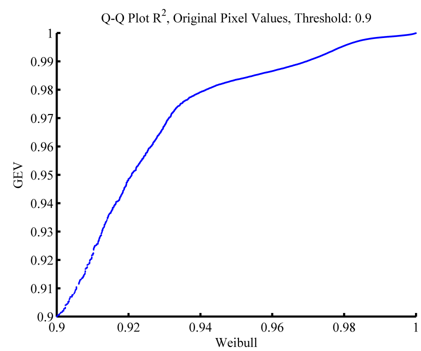
First we illustrate the quality of the fittings for the Weibull and the GEV-case and give an overview over some results regarding the goodness-of-fit. We used the *coefficient of determination* (denoted by R^2) as g.o.f. measure which measures the similarity between the empirical cdf and the model cdf. It has a maximum value of one (see [20]). We found that the GEV-distributions gave higher fitting values than the corresponding Weibull distributions and that the fittings were slightly better for the log-transformed images.

For some of the images either the Weibull or the GEV distributions (or both) did not fit the data very well. In Table 1 we show the mean values of the R^2 values taken over all images where the mininum value was greater 0.5 (first line) or greater than 0.9 (second line). We see that the values for the GEV-fittings is slightly higher for both the results computed from the original pixel values (columns three and four) and from the values computed from the logarithmic pixel values (columns five and six). We see also that the vast majority of slices (and certainly all slices with meaningfull content) fit these distributions (column two).

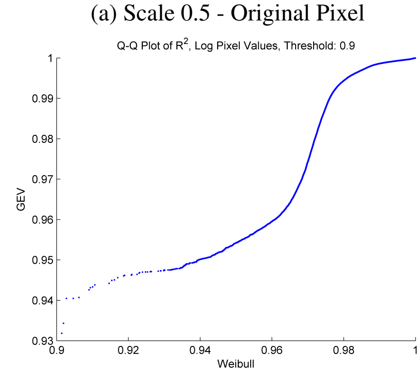
T.	Images	Wbl.	GEV	LogWbl.	LogGEV
0.5	51323	0.9776	0.9856	0.9774	0.9888
0.9	49148	0.9824	0.9947	0.9820	0.9891

Table 1: Mean R^2 Values

More detailed information of the relation between the values for the Weibull and the GEV-fittings can be seen in the quantile-quantile plots in Figure 2.



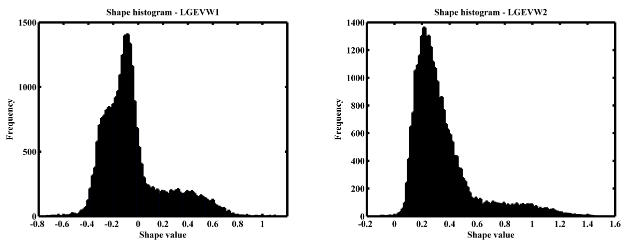
(a) Scale 0.5 - Original Pixel



(b) Scale 0.5 - Log Pixel

Fig. 2: Quantile-quantile plot of g.o.f. R^2-Values

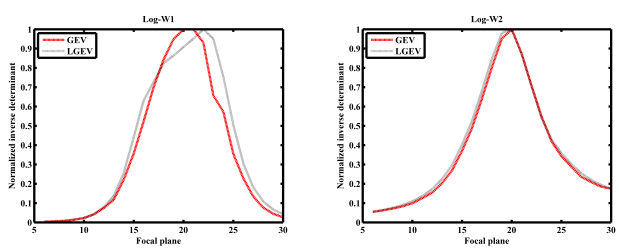
The GEV distribution comes in three different types, depending on the value of the shape parameter ξ. For the experiments with the logarithmic pixel values we illustrate the distribution of the shape parameters obtained for the two image types W1 and W2. The result in Fig 3 shows that almost all distributions for image type W2 have positive shape parameters whereas a majority of the distributions fitted to the W1 images have negative shape values.



(a) Image Type W1 (b) Image Type W2

Fig. 3: Histograms of Shape Parameter for Log. Pixel Values

An analysis of the Fisher matrices for the 3-parameter Weibull fittings (using Matlabs wblfit function) shows that many distributions are actually located on a two-dimensional submanifold. This confirms earlier observations that it is sufficient to consider the two-parameter Weibull distribution which ignores the possible influence of the location parameter. In Figure 4b we show the mean values of the (normed) inverse of the determinant of the Fisher matrix of the GEV distributions as a function of the focus setting. Here we used the Matlab gevfit function to compute the maximum-likelihood estimates of the parameters. We see a clear difference between the images based on the two different stainings but all of them peak in the focus interval z=16 and z=21 which is an acceptable focus range according to the benchmark description.



(a) Image Type W1 (b) Image Type W2

Fig. 4: Inverse of the Determinant Value of the GEV-Fisher Matrix as a Sharpness Function

Figure 4 shows that the determinant of the Fisher matrix provides a measure of the sharpness of an image. It is however static and does not take into account the relations between the current image and the neighboring images in the focus sequence. Intuitively one can guess that the best focus position should be a critical point where the increased sharpening changes to an increased blurring. The curve describing the changes of the distributions in the manifold should therefore have a critical point at the best focal position too. We can therefore use the information about the gradient of the determinant to improve the focusing. As an illustration we computed the cosine of the angle between the gradient vectors of the determinant sharpness function from two consecutive focal images. The angle is computed in the geometry given by the Fisher matrix. When these vectors point into the same direction (focusing or blurring) the angle should be small and the cosine large. For the critical point the cosine should be small. The mean value of the cosine values computed over all images of type W1 computed from the log-pixel images is shown in Fig. 5. We can see a clear local minimum for focus position 16 which is the ground truth position.

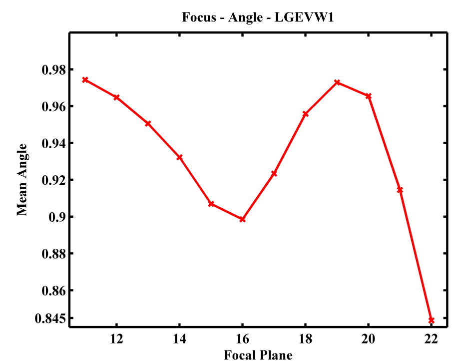


Fig. 5: Cos Angle LogGEVW1

4. SUMMARY AND CONCLUSIONS

We introduced GEV-distributions in the analysis of visual descriptors of the edge-magnitude type. We showed that for large classes of images they provide very good statistical fittings. Since they are related to the exponential distribution we were also able to derive analytical expressions of the Fisher matrix and functions derived from it. In particular we illustrated how to use the gradient of the determinant of the Fisher matrix as an additional sharpness indicator.

No attempts were made to develop efficient implementations of the processing steps and the computation times are therefore high but they can certainly be improved using optimization methods similar to those described in [13].

Finally we want to point out that the results described are derived using three general principles: The filter functions are based on the symmetry properties of the sensor array, the selection of the distributions is motivated by their characteristic limit and stability properties and the determinant of the Fisher matrix as a sharpness function is based on the geometric property of the volume. All of them are natural choices in the general context in which they are introduced and none of them is designed for the specific task of designing an autofocus algorithm. This framework should therefore also be useful in other applications besides analyzing the focusing properties of imaging systems.

5. REFERENCES

[1] E.P. Simoncelli, "Statistical models for images: compression, restoration and synthesis," in *Thirty-First Asilomar Conf. on Signals, Systems & Computers*, 1997, vol. 1, pp. 673–678.

[2] Y. Jia and T. Darrell, "Heavy-tailed distances for gradient based image descriptors," in *Advances in Neural Information Systems*, Granada, 2011, pp. 1–9.

[3] J.M. Geusebroek and A.W.M. Smeulders, "A six-stimulus theory for stochastic texture," *Int. J. Computer Vision*, vol. 62, no. 1-2, pp. 7–16, 2005.

[4] H.S. Scholte, S. Ghebreab, L. Waldorp, A.W.M. Smeulders, and V.A.F. Lamme, "Brain responses strongly correlate with Weibull image statistics when processing natural images," *J. Vision*, vol. 9, no. 4, 2009.

[5] V. Yanulevskaya and J.M. Geusebroek, "Significance of the Weibull distribution and its sub-models in natural image statistics," in *Int. Conf. Computer Vision Theory and Applications*, 2009, pp. 355–362.

[6] V. Zografos and R. Lenz, "Spatio-chromatic image content descriptors and their analysis using extreme value theory," in *Image Analysis*, *LNCS-6688*, pp. 579–591, Springer, 2011.

[7] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE TPAMI*, vol. 33, no. 4, pp. 687–698, 2011.

[8] R.E. Kass and P.W. Vos, *Geometrical foundations of asymptotic inference*, Wiley, New York, 1997.

[9] S. Amari and H. Nagaoka, *Methods of information geometry*, American Mathematical Society, 2000.

[10] H. Rinne, *The Weibull Distribution: A Handbook*, CRC Press, 2008.

[11] P. Prescott and A.T. Walden, "Maximum likelihood estimation of the parameters of the generalized extreme-value distribution," *Biometrika*, vol. 67, no. 3, pp. 723–724, 1980.

[12] P.J. Olver, *Applications of Lie groups to differential equations*, Springer, New York, 1993.

[13] V. Zografos, R. Lenz, and M. Felsberg, "The Weibull manifold in low-level image processing: An application to automatic image focusing," *Image and Vision Computing*, vol. 31, no. 5, pp. 401–417, 2013.

[14] M. A. Bray, A. N. Fraser, T. P. Hasaka, and A. E. Carpenter, "Workflow and metrics for image quality control in large-scale high-content screens," *J. Biomolecular Screening*, vol. 17, no. 2, pp. 266–274, 2012.

[15] R. Lenz, "Investigation of receptive fields using representations of dihedral groups," *J. Vis. Comm. Im. Rep.*, vol. 6, no. 3, pp. 209–227, 1995.

[16] R. Lenz, V. Zografos, and M. Solli, "Dihedral color filtering," in *Advanced Color Image Processing and Analysis*, Springer, 2013.

[17] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int J Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[18] M.R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York, 1983.

[19] J. Beirlant, *Statistics of extremes : theory and applications*, Wiley, Hoboken, N.J., 2004.

[20] E.L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statist. Assn.*, vol. 53, pp. 457–481, 1958.