Characterization of SURF Interest Point Distribution for Visual Processing in Sensor Networks

Muhammad Altamash Khan, György Dán, Viktoria Fodor
ACCESS Linnaeus Center, School of Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden
{khan6,gyuri,vfodor}@ee.kth.se

***Abstract*—We study the statistical characteristics of SURF interest points and descriptors, with the aim of supporting the design of distributed processing across sensor nodes in a resource constrained visual sensor network. We consider a sensor network with a single camera node and four schemes of delegating processing tasks to the sensor nodes. We discuss the potential and the challenges of the different schemes in light of the results of the statistical analysis. Our results show that the distribution of the number of interest points per image exhibits a heavy tail. The interest point locations are almost uniformly distributed along the axes of the images, but their X and Y coordinates are slightly correlated. Most interest points are found in the lowest octave layers, and the number of interest points decreases exponentially with scale. Our analysis suggests that for a wireless broadcast channel delegating subareas of images to processing nodes would lead to a more even allocation than delegating by octave layers. For directional wireless channels the efficiency can be significantly improved by performing some of the feature extraction tasks at the camera node.**

***Index Terms*—SURF; visual sensor network; interest point distribution**

I. Introduction

In visual sensor networks (VSNs), images or video sequences captured by low cost cameras need to be processed and transmitted to a central node or image database to perform visual analysis tasks, like image retrieval or video data mining [1]. Visual analysis is often based on the detection and extraction of visual features, that can characterize the image. The detected set of features should allow high performance visual analysis, for example, in the case of image retrieval high precision and recall. Several feature extraction techniques have been proposed in the last decades, providing scale and rotation invariance [2], [3], [4], [5]. In this work we consider one of the most popular ones, Speed-up Robust Features (SURF) [3].

In general, feature extraction techniques consider the pixel data of an image, find a set of interest points and extract the related feature descriptors. The feature descriptors are then used to e.g., identify images in a database. To perform these tasks in a sensor network environment is challenging, since the tasks are both computationally demanding and require the transmission of a large amount of information, the raw pixel data or the generated feature descriptors. To decrease the amount of data to be transmitted and thus the system latency, [6] proposes a progressive transmission scheme, that terminates the transmission of new descriptors if the image is retrieved. In [7] a distributed solution is presented, where descriptors of objects are stored at the nodes in the VSN, and new queries are routed to these nodes for object matching. In [8] the number of considered interest points and the quantization level of the descriptors are jointly optimized to maximize the accuracy of the recognition, subject to energy and bandwidth constraints.

In this work we investigate how to best distribute the tasks of interest point detection and descriptor extraction across the nodes of the VSN such that the VSN resources are used efficiently and the computational and transmission load are balanced. We define four strategies for off-loading the camera node and present a detailed statistical evaluation of the number, location and scale of SURF interest points. Based on our results we discuss the expected performance of the different off-loading schemes. To the best of our knowledge this work is the first to provide a statistical characterization of SURF interest points based on a large image data set.

The rest of the paper is organized as follows. In Section II we give a background on SURF, describe the VSN and the four off-loading strategies. In Section III we provide a detailed statistical analysis of the SURF interest points. Finally, in Section IV we discuss the results and conclude our work.

II. Background and System model

We consider a VSN consisting of a camera node C, a set $\mathcal{P}$ of processing nodes and a data storage node S. The processing capability and energy resources of all nodes are limited as well as the communication bandwidth. The transmission channel from C to $\mathcal{P}$ may be broadcast, or may consist of directional links. The goal of the VSN is to extract SURF descriptors in the images captured at C and to transmit them to S. Fig. 1 illustrates the considered VSN scenario.

The computation of SURF descriptors involves performing three main processing steps. Given an image i, the first step involves the *detection* of interest points.

To provide scale invariance, SURF calculates blob response maps at different scales, by convolving the integral image with Gaussian box filters of increasing size. The standard deviation σ of the Gaussian filter represents the scale. The scale space is divided into octaves, and each octave is divided into octave layers. Responses at the neighboring octave layers within an octave are compared to find the interest points, and the

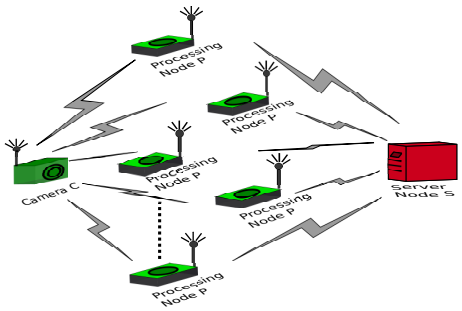


Fig. 1: Considered visual sensor network scenario consisting of a camera node, processing nodes and a data storage server node.

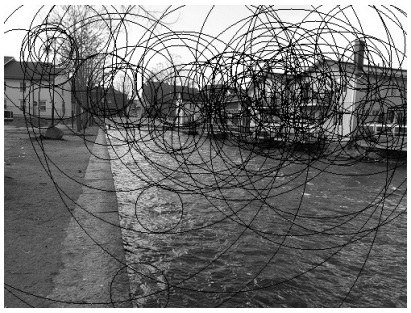


Fig. 2: Circular interest areas in a sample image for detection threshold $\Theta = 4900$.

scale of an interest point is determined through interpolation with the neighboring layers. The detection threshold value Θ determines the required contrast to the neighboring areas to detect a blob response. The parameters of the interest point detection process are the number of octaves and octave layers and the threshold value, and these parameters need to be selected based on the requirements of the visual task [3]. The number K_i of interest points in image i is unknown until interest point detection is done for all octave parameters of interest.

Every detected interest point k in image i is characterized by its location $(x,y)_{i,k}$ and its scale $\sigma_{i,k}$. These two parameters identify a circular interest area $A_{i,k}$ of diameter $\sqrt{2}\cdot 20\sigma_{i,k}$ within the image. Fig 2 shows the circular interest areas in a sample image. The second step involves *orientation* identification, when a rectangular interest area $R_{i,k}$ of side length $20\sigma_{i,k}$ is selected within the circular interest area $A_{i,k}$ for every interest point. The last step is the *extraction* of the SURF descriptor $D_{i,k}$ based on $R_{i,k}$ for every interest point k in image i. Every descriptor has the same size, unless coding with variable accuracy is used. To accelerate the extraction process and to extend the lifetime of the VSN, some or all processing steps have to be delegated from C to $\mathcal{P}$.

The *delegation of the interest point detection* can be done in three ways. C can delegate to a node $P \in \mathcal{P}$ an area $Z_{i,j}$ of image i to be processed, which we call *area-split*. The areas need to overlap in order to be able to detect all interest points, which might require redundant data transmission if the wireless links are directional. The width of the overlap is determined by the size of the largest *interest area*, and is therefore $\sqrt{2}\cdot 10$ times the expected largest scale. With area-split the number of interest points detected by node P depends on the spatial distribution of the interest points.

Alternatively, C can delegate to a node $P \in \mathcal{P}$ the SURF octave parameters (effectively the scales) to be used for interest point detection, which we call *scale-split*. In this case all nodes need to receive the entire image, and the number of interest points detected by P depends on the distribution of the interest points across scales. Finally, the delegation of interest point detection can be based on both area and scale.

The *delegation of orientation identification* requires the pixel data for $A_{i,k}$ to be transmitted, and the *delegation of descriptor extraction* requires the pixel data for $R_{i,k}$ to be transmitted. The interest areas for different interest points might overlap, and this overlap can potentially be used to decrease the amount of data transmitted between the nodes for the delegation of the tasks.

The delegation of the processing steps affects the use of the computational and communication resources of the VSN nodes: the data transmission from C to $\mathcal{P}$, the computational load of C and the nodes in $\mathcal{P}$, and the data to be transmitted from $\mathcal{P}$ to S. These three are strongly coupled, and therefore the delegation needs to be optimized. We consider the following alternatives of delegating processing steps in order to off-load the camera node C:

- *No Detection / No Extraction (ND/NE)*: The camera node performs neither detection nor extraction. Thus, the entire image needs to be sent to the processing nodes, and area-, scale- or hybrid split can be used. In the case of scale-split the delegation has to be based on the parameters of the detection algorithm, that is the octaves and octave layers, and each of the processing nodes should detect interest points with distinct input parameters to avoid redundant detection. Redundant processing can be avoided by delegating different octaves to the processing nodes.
- *Partial Detection / Partial Extraction (PD/PE)*: The camera node detects some of the interest points and extracts the related descriptors. Through detecting and extracting interest points at high octaves, that is, large scales, the redundancy of data transmission under area-split can be decreased.
- *Complete Detection / No Extraction (CD/NE)*: The camera node detects all interest points and delegates the descriptor extraction to the processing nodes. Given the location and scale of all interest points, the camera node needs only transmit the pixel data for each interest area $A_{i,k}$. If the camera node calculates orientation as well, only the pixel data of $R_{i,k}$ is needed, which can decrease the number of pixels to be transmitted by up to one third. At the same time the vector of interest point locations and scales needs to be transmitted. The balancing of the computation among processing nodes can be done based on the known location and scale parameters and becomes trivial if a broadcast link is considered. In the case of directed links, the communication cost needs to

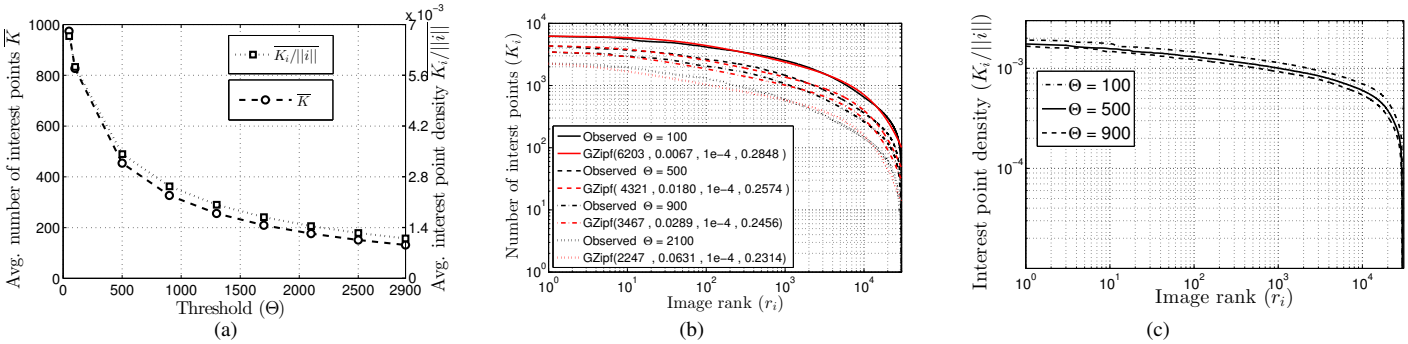


Fig. 3: (a) Average number of detected interest points and average interest point density vs. detection threshold. (b) Rank statistics of the number of interest points for detection thresholds $\Theta = 100$, 500, 900 and 2100, and best-fit generalized Zipf distribution. (c) Rank statistics of the interest point density for detection thresholds $\Theta = 100$, 500, and 900.

be balanced as well. In this case the problem is related to graph clustering and is known to be NP-complete.

- *Complete Detection / Partial Extraction (CD/PE)*: The camera node detects all interest points and extracts some of the descriptors. The set of interest points for which descriptors are extracted at the source should be chosen to minimize the remaining image pixels to be transmitted to $\mathcal{P}$. In this paper we consider a simple heuristic, which consists of processing interest points with largest scales at the source. This heuristic would be optimal if interest areas would not overlap.

Which of these options can achieve the best performance depends on the computational and transmission resources and on the locations and scales of the interest points, which is unknown a priori. In the next section we evaluate statistical properties of the interest points, considering their number, location and scale-size distribution, and evaluate the expected computational and transmission gains of the above delegation schemes.

III. Statistical Characterization

We processed the 30607 images of the Caltech-256 object category data set [9] using the SURF implementation of OpenCV [10]. The images in this dataset are of variable size, ranging from $5 \cdot 10^3$ to $12 \cdot 10^6$ pixels. For interest point detection we used the default SURF parameters, 4 octaves and 2 octave layers within each octave, and we considered various detection threshold values, from 100 to 2900. The default threshold value was $\Theta = 500$, unless otherwise noted.

A. Number of interest points

The number K_i of interest points in an image affects both the load of the processing nodes and the amount data to be transmitted to the server node S. Therefore we start with the evaluation of the distribution of the number of interest points over all images.

Figure 3a shows the mean number of detected interest points $\bar{K}$ as a function of the detection threshold Θ. The number of interest points decreases sharply for low threshold values and by increasing the detection threshold value significantly the number of interest points decreases by an order of magnitude, although potentially at the price of decreased recall and precision. At the same time, the rate of decrease diminishes as the threshold Θ increases. We see a similar trend in the results for the average interest point density, which is the average of the number of detected interest points normalized by the image size, (i.e., $K_i/||i||$). Therefore, we can conclude, that increasing an already high threshold value will not significantly decrease the computational and transmission costs in the VSN.

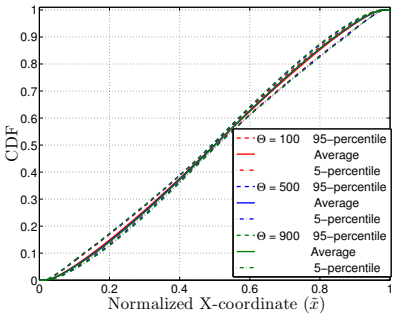
In Figure 3b we show the rank statistic of the number of detected interest points at four different detection threshold values for all images. The rank statistic is obtained by showing the number of interest points K_i in decreasing order for all images. Thus, the leftmost value corresponds to $\max_i K_i$ and the rightmost value to $\min_i K_i$ The number of interest points varies by up to two orders of magnitude between images, from a couple of tens to a couple of thousands. Nevertheless, the shape of the curves does not change significantly when increasing the detection threshold Θ.

In order to better understand the distribution of the number of interest points we fitted the generalized Zipf law to the data [11]. According to the generalized Zipf law the frequency of an object with rank r is given by

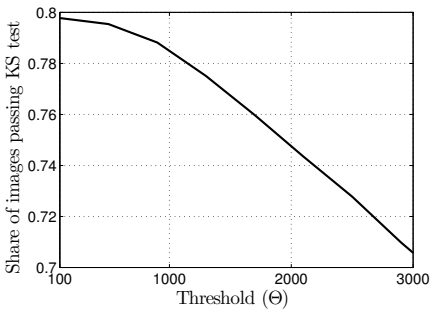
$$f_{GZipf(f_1,\lambda,\mu,\theta)}(r) = \frac{f_1}{\left[1 - \frac{\lambda}{\mu} + \frac{\lambda}{\mu}\, e^{\frac{1}{\theta}\mu r}\right]^{\theta}}. \qquad (1)$$

The generalized Zipf law exhibits an initial plateau, a power-law trunk and an exponential cut-off, and was used to model proportional voting processes and content popularity in the Internet [11], [12]. Figure 3b shows that the fitting is rather accurate, and suggests that there might be a power-law scaling in the number of interest points per image.

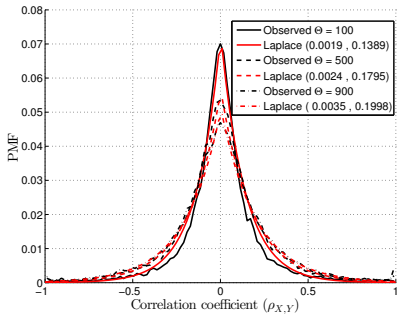
The number of interest points in an image can, of course, depend on the image size, and thus the large difference between the number interest points per image could be partly due to the differing sizes of the images in the data set. Figure 3c shows the rank statistics of the interest point density (i.e., $K_i/||i||$) for three detection threshold values for all images. The variability



(a) CDF of normalized X-coordinates, average, 5 and 95 percentiles. The distributions are close to uniform for the majority of the images.



(b) Fraction of images that pass the two-sample K-S test at significance level 0.05 between the distributions of the normalized X- and Y-coordinates.



(c) PMF of the correlation coefficients $\rho^i_{X,Y}$, with best fit Laplace distributions. Big correlation is more common for large thresholds.

Fig. 4: Characteristics of the location distribution of interest points for various detection threshold values.

of the interest point density is lower than that of the number of interest points, but it is still up to one order of magnitude. Consequently, one can expect a high variability in the number of interest points detected in an image. This in turn makes it necessary to adapt a potential off-loading scheme and the choice of SURF parameters depending on the image content.

B. Spatial distribution of interest points

The spatial distribution of the interest points in an image plays an important role in the efficiency of *area-split*. To investigate the spatial distribution of interest points, let us define $\tilde{x}_{i,k}$ as the normalized X-coordinate of interest point k in image i, i.e., $x_{i,k}$ normalized with the width of image i. Furthermore, we define $n_{i,\tilde{x}}$ to be the number of interest points with normalized X-coordinate less than or equal to $\tilde{x}$ in image i, i.e., $n_{i,\tilde{x}} = |\{k|\tilde{x}_{i,k} \leq \tilde{x}\}|$, and let $F_i(\tilde{x}) = \frac{n_{i,\tilde{x}}}{K_i}$.

Figure 4a shows the average, and the 5 and 95 percentiles of $F_i(\tilde{x})$ for threshold levels $\Theta = 100$, 500 and 900. The distribution is close to uniform apart from the border regions, independently from the detection threshold. Furthermore, the 5 and 95 percentiles are very close to the mean value, which means that the spatial distribution of the interest points is close to uniform in the individual images too. Also, the distribution is fairly insensitive to the threshold value. We found identical results for the distribution of the normalized Y-coordinates. We performed a two-sample K-S test to investigate whether the distribution of the normalized X-coordinates is the same as that of the normalized X-coordinates. The results in Fig. 4b show that the fraction of images for which this hypothesis cannot be rejected at a significance level of 0.05 is generally high, but decreases almost linearly with the threshold Θ, thus as the number of interest points decreases. This phenomenon could potentially be explained by a correlation between the coordinates when there are few interest points.

To investigate the correlation between the X and Y-coordinates, Figure 4c shows the probability mass function of the correlation coefficients $\rho^i_{X,Y}$ of the X and Y-coordinates of the interest points over all images i, for three threshold values. The figure also shows the best fit Laplace distributions, with their location and scale parameters. The Laplace distribution provides the best fit for $\Theta = 900$, but this is also the threshold value for which the correlations are highest, which supports that correlation could be the reason for images not passing the two-sample K-S test. The value of the location parameter of the Laplace distributions, around 10^{-3} for all three threshold values, shows that the average correlation coefficient $\overline{\rho}_{X,Y}$ over all images is close to zero. Furthermore, the good fit of the Laplace distribution implies that the distribution of the absolute value of the correlation coefficents is close to exponential, i.e., large correlations are rare. The prevalence of small correlation values could be a sign of that the X and Y-coordinates of the interest points tend to be independent.

To verify this hypothesis (H_0) we performed Pearson's χ^2 test for independence between the X and Y-coordinates for every image. We divided the X and Y-axis into 6 intervals each, and used the resulting 36 rectangles and the marginal distributions along the X and Y-axis for the χ^2 test. The distribution of the χ^2 values for various detection thresholds are shown in Figure 5 together with the critical χ^2 values for a significance level of 0.05 and 0.01. The results show that between 20 to 30 percent of the images pass the test at a significance level of 0.05, i.e., the coordinates of the interest points might be independent.

Based on the spatial distribution of the interest points we can conclude that if area-split is used for the *ND/NE* and for the *PD/PE* schemes then the average load on the processing nodes can be made close to uniform even without a priori knowledge of interest point locations. The low correlation between the coordinates and the χ^2 test results suggest that area-split can be optimized by optimizing independently the horizontal and the vertical cuts based on the distribution shown in Figure 4a. Thus, area-split can be a good solution in the case of broadcast transmission from C to $\mathcal{P}$.

C. Scale size distribution

We now turn to scale-split and investigate the distribution of the number of interest points according to the octave layers they were detected at and across the scales after interpolation.

1) Interest points per octave: If scale-split is applied in the *ND/NE* or the *PD/PE* schemes then the delegation needs to

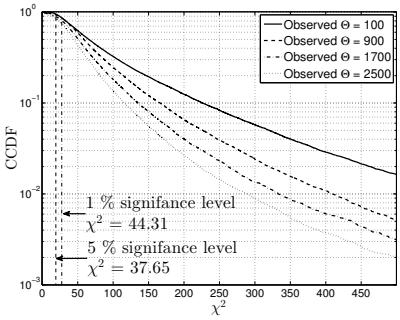
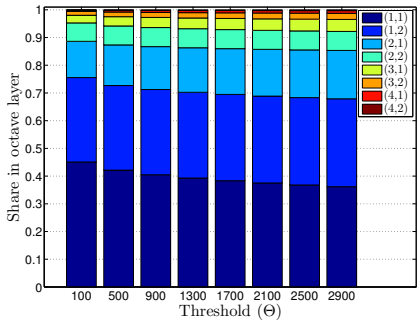
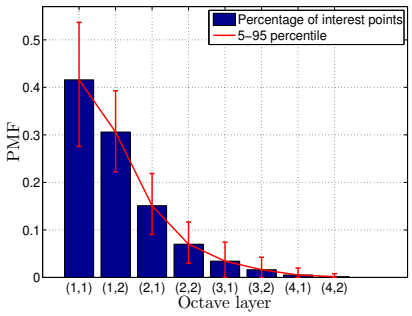


Fig. 5: CCDF of χ^2 values obtained from independence test with 0.01 and 0.05 significance levels marked. About 20 percent of the images pass the test.



(a) Share of interest points per octave layer vs. detection threshold. Most interest points are detected in the first octaves.



(b) Interest point distribution for $\Theta = 500$. Mean, 5 and 95 percentiles. The distribution is skewed, and varies significantly between images.

Fig. 6: Distribution of interest points per octave layer.

be based on the octaves and octave layers, to avoid redundant processing and detection. Therefore, we first evaluate the distribution of the number of interest points among the octaves and octave layers. As we consider the default parameters of 4 octaves and 2 central octave layers within each octave, the interest point detection is performed in 8 rounds, with different filter sizes.

Figure 6a shows the mean value of the share of interest points detected at the different octave layers for various detection threshold values. We denote the octave layers by the (octave, central octave layer) parameter pair. The figure confirms that the majority of interest points are detected in the first octaves [3]. Nevertheless, we see a general trend that the percentage of interest points detected at higher octaves increases as the threshold Θ is increased. Extreme cases were reported for a limited set of images for very high threshold values in [13].

Figure 6b shows the probability mass function of the interest point distribution across the octave layers, averaged for all images, and the 5 and 95 percentile bar, for the default $\Theta = 500$. The distribution is very skewed; processing interest points at the 2 or 4 highest octave layers, that is, layers (4,1),(4,2) and (3,1),(3,2),(4,1),(4,2) respectively, would result in 0.3% and 7% of the total processing. The skewed distribution would make the delegation of processing based on the octave layers very unbalanced, and therefore scale-split based on octave layers would have to be combined with area-split, rotating octave assignment, or a subsequent re-assignment of descriptor extraction to achieve similar processing and transmission loads at $\mathcal{P}$. Considering the *ND/NE* and *PD/PE* schemes, the allocation of octaves and layers to processing nodes would have to be done a priori, possibly based on the average values shown in Figure 6b. Unfortunately, the 5 to 95 percentile ranges are rather wide: the actual computation load for a particular image may differ by more that 50% from the average for more than 10% of the images.

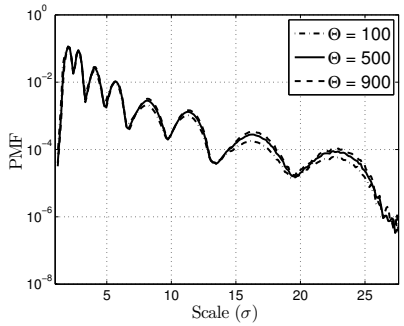
2) Interest points per scale: In the following we evaluate the distribution of the scales of the interest points. Figure 7a shows the averaged PMF of the scale size distributions of all images for three threshold levels on a logarithmic scale. The 8 distinct peaks reflect the 8 octave layers. We can see that the PMF curves loosely follow the exponential distribution, which means that the vast majority of interest points has a relatively small scale. We also observe that the effect of the threshold value on the distribution is very small.

We next look at the ratio of interest points falling below a given scale for all considered images. Let $n_{i,\sigma}$ be the number of interest points with scale less than or equal to σ in image i, i.e., $n_{i,\sigma} = |\{k|\sigma_{i,k} \leq \sigma\}|$, and let $F_i(\sigma) = \frac{n_{i,\sigma}}{K_i}$. Figure 7b shows the average, and the 5 and 95 percentiles of $F_i(\sigma)$ for threshold level $\Theta = 500$. The figure confirms that the majority of the interest points are detected at low scales, but the scale size distribution may differ significantly from image to image. Similar results with different threshold values show that the difference from the average increases slightly as the detection threshold Θ is increased. These results again confirm that the variability of the load of the processing nodes under scale-split could be significant.

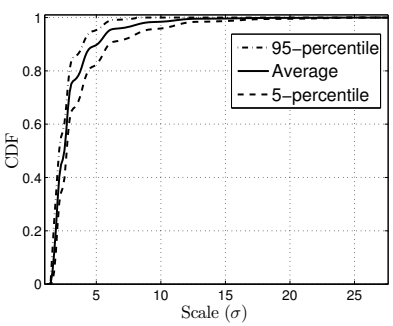
Comparing the statistics of the interest point coordinates and that of the scales we can conclude that for *ND/NE* and for *PD/NE* area-split would lead to a more balanced load distribution in the processing nodes than scale-split, and therefore area-split would be preferable if there is a broadcast channel connecting C and $\mathcal{P}$. In the case of *CD/NE* and *CD/PE* the scale-size distribution is known at C and it could be used for balanced delegation of descriptor extraction among the nodes in $\mathcal{P}$.

D. Transmission redundancy

So far we considered a broadcast link between C and $\mathcal{P}$. Let us now evaluate the potential redundancy of area-split under *ND/NE* and *PD/PE* if the links are directional. We quantify the redundancy by the ratio of the sum of the areas to be transmitted and the actual size of the image, normalized by the number of areas n_Z, i.e., $(\sum_j ||Z_{i,j}||/||i||)/n_Z$. For example, a redundancy value of 1 means that the entire image has to



(a) PMF of scales for three detection thresholds. The PMF curves loosely follow an exponential distribution.



(b) CDF of scales at $\Theta = 500$. Most interest points have a small scale, but differences can be large between images.

Fig. 7: Distribution of interest point scales $\sigma_{i,k}$.

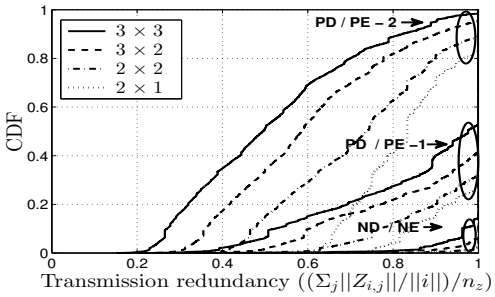


Fig. 8: CDF of normalized transmission redundancy under area-split for *ND/NE* and for *PD/PE-n* for $n = 1, 2$ octaves processed at *C*.

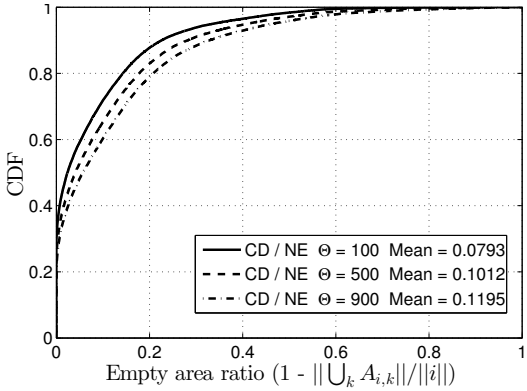
be transmitted n_Z times.

Figure 8 shows the CDF of the redundancy when the images are divided into 2×1, 2×2, 3×2 and 3×3 equal sized areas. Results are shown for *ND/NE* and for *PD/PE-n* for $n = 1, 2$, when n highest octaves are processed at *C*. That is, under *PD/PE-1* octave layers (4,1) and (4,2) are processed at *C*, while under *PD/PE-2* octave layers (3,1), (3,2),(4,1) and (4,2) are processed at *C*. For *ND/NE* the maximum scale that needs to be detected under the default 4 octaves and 2 octave layers is $\sigma = 27.6$, for *PD/PE-n* it is $\sigma = 14.26$ and 7.06 for $n = 1$ and $n = 2$, respectively. The results show that under *ND/NE* there is a very high probability, between 0.7 to 0.9, that the entire image needs to be transmitted for every individual area $Z_{i,j}$. For example, for the 3×2 area split the probability that the entire image needs to be transmitted six times is almost 0.8. With *PD/PE-n* the redundancy can be significantly reduced compared to *ND/NE*; the average reduction is almost 60% for the 3×3 scenario. Consequently, *ND/NE* with area-split is not a favorable scheme if the *C* to $\mathcal{P}$ links are directional and the transmission bandwidth or energy are scarce resources. In that case *PD/PE* has to be used, and needs to be adapted to the image size $||i||$.

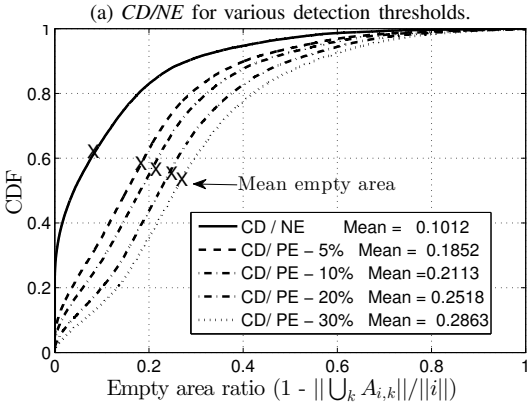
Reconciling these results with those for the distribution of interest points among octaves in Figure 6b shows that, although processing the 2 (*PD/PE-1*) or 4 (*PD/PE-2*) highest octave layers would lead to 0.3% and 7% of the processing to be done at the camera node, respectively, doing so would decrease the average transmission redundancy by 20%-60%, depending on the area-split configuration. That is, partial detection and extraction at the camera node could be very beneficial to area-split if the links are directional.

E. Empty area distribution

Let us finally consider the *CD/NE* and the *CD/PE* schemes, that is, when the camera node *C* performs the detection of *all* interest points. As in this case the location and the scale of the interest points are known, only the interest areas around the interest points, that is, the discs with radius $\sqrt{2}\cdot10\sigma$ need to be transmitted to the nodes in $\mathcal{P}$ for feature extraction. We refer to the area of an image not covered by any interest area



(a) *CD/NE* for various detection thresholds.



(b) *CD/PE* and detection threshold $\Theta = 500$

Fig. 9: Empty area ratio under *CD/NE* and *CD/PE*.

as *empty area*. As the ratio of the empty area increases, the transmission cost is decreased.

1) Empty area distribution under CD/NE*:* Figure 9a shows the ratio of the empty area defined as the area of the image not covered by any interest area, i.e., $1 - ||\bigcup_k A_{i,k}||/||i||$ for 500 images for various threshold values. The ratio of the empty area turns out to be very small for all cases. It increases slightly with increased detection threshold because less interest points are detected. For the default $\Theta = 500$, the average achievable gain of not transmitting empty areas from the camera node to the processing nodes is slightly below 10%. At the same time,

ca. 30% of the images do not have empty area at all. Thus, *CD/NE* might not significantly reduce the amount of data to be transmitted from the camera node to the processing nodes.

2) Empty area distribution under CD/PE*:* With partial feature extraction performed at the camera node, that is, under *CD/PE*, the empty area can be increased since the interest areas that correspond to interest points processed at the camera node *C* do not need to be transmitted, only the extracted descriptors. In principle the set of interest points extracted at the camera node could be chosen such as to maximize the empty area, but doing so requires combinatorial optimization due to the overlaps between interest areas. We therefore applied a heuristic to choose the interest points to be extracted at *C*: a fraction of the interest points, starting from the highest scale, are processed locally at *C*. Figure 9b compares the CDF of the ratio of the empty area under *CD/NE* and *CD/PE* when 5% to 30% of feature descriptors are extracted locally at *C*. As we see, extracting 10% of the descriptors at *C* already doubles the empty area on average, and the probability that an image does not have empty area at all gets close to zero. Consequently, if the transmission resources in the VSN are limited, it may be beneficial to combine local interest point detection with partial feature extraction.

IV. Conclusion

We provided a statistical characterization of SURF interest points based on a public image database with the aim of getting insights into the potential performance of different processing offloading schemes in a visual sensor network. Our results show that area-split can be an efficient solution if the camera to processing node communication is broadcast. In the case of directional links area-split would involve high redundancy, but the redundancy could be decreased through partial detection and extraction at the camera node. Scale-split without complete detection at the camera node would lead to highly non-uniform load among the processing nodes, unless dynamic load balancing is implemented. The complete detection of interest points, togehter with partial descriptor extraction, could be used to decrease the total amount of data to be transmitted to the processing nodes, by avoiding the transmission of pixels not covered by interest areas.

References

[1] A. Marcus and O. Marques, "An eye on visual sensor networks," *IEEE Potentials*, vol. 31, no. 2, pp. 38 –43, Apr 2012.

[2] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. of IEEE ICCV*, 1999, pp. 1150–1157.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008.

[4] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. of ECCV*, 2010.

[5] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. of IEEE ICCV*, 2011.

[6] V. R. Chandrasekhar, S. S. Tsai, G. Takacs, D. M. Chen, N.-M. Cheung, Y. Reznik, R. Vedantham, R. Grzeszczuk, and B. Girod, "Low latency image retrieval with progressive transmission of chog descriptors," in *Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing*, 2010.

[7] V. Sulic, J. Pers, M. Kristan, and S. Kovacic, "Efficient feature distribution for object matching in visual sensor networks," *IEEE Trans. on Circ. Syst. Video Techn. (CSVT)*, vol. 21, no. 7, pp. 903–916, 2011.

[8] A. Redondi, M. Cesana, and M. Tagliasacchi, "Rate-accuracy optimization in visual wireless sensor networks," in *Prof. of IEEE ICIP*, 2012.

[9] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007. [Online]. Available: http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001

[10] "OpenCV." [Online]. Available: http://opencv.org/

[11] G. Dán and N. Carlsson, "Power-law revisited: A large scale measurement study of p2p content popularity," in *Proc. of 9th International Workshop on Peer-to-Peer Systems (IPTPS '10)*, Apr 2010.

[12] M. L. Lyra, U. M. S. Costa, R. N. C. Filho, and J. S. Andrade, "Generalized zipf's law in proportional voting processes," *Europhys. Lett.*, vol. 62, no. 1, pp. 131–137, 2003.

[13] S. Ehsan, N. Kanwal, E. Bostanci, A. F. Clark, and K. D. McDonald-Maier, "Analysis of interest point distribution in SURF octaves," in *3rd International Conference on Machine Vision (ICMV 2010)*, Apr 2010.