On the Assessment of Segmentation Methods for Images of Mosaics

Gianfranco Fenu[1], Nikita Jain[2], Eric Medvet[1], Felice Andrea Pellegrino[1] and Myriam Pilutti Namer[3]
[1]*Department of Engineering and Architecture, University of Trieste, Trieste, Italy*
[2]*Department of Software Engineering, Delhi Technological University, Delhi, India*
[3]*Department of Humanities, Ca' Foscari University of Venice, Venice, Italy*

Keywords: Image Segmentation, Superpixel, Comparative Evaluation, Cultural Heritage Preservation.

Abstract: The present paper deals with automatic segmentation of mosaics, whose aim is obtaining a digital representation of the mosaic where the shape of each tile is recovered. This is an important step, for instance, for preserving ancient mosaics. By using a ground-truth consisting of a set of manually annotated mosaics, we objectively compare the performance of some existing recent segmentation methods, based on a simple error metric taking into account precision, recall and the error on the number of tiles. Moreover, we introduce some mosaic-specific hardness estimators (namely some indexes of how difficult is the task of segmenting a particular mosaic image). The results show that the only segmentation algorithm specifically designed for mosaics performs better than the general purpose algorithms. However, the problem of segmentation of mosaics appears still partially unresolved and further work is needed for exploiting the specificity of mosaics in designing new segmentation algorithms.

1 INTRODUCTION AND RELATED WORK

Mosaics are artworks reproducing decorative images or patterns out of small colored tiles (called *tesserae* or *tessellae*). Ancient mosaics represent a relevant part of the cultural heritage of many countries and digital imaging may provide tools for their preservation, distant consultation, reconstruction, anastylosis (Benyoussef and Derrode, 2011). Traditionally, for preservation purposes, ancient mosaics are manually acquired by tracing the contour of each tile over a semi-transparent paper superimposed to the mosaic. Needless to say, this is a time-consuming task but due to the specificity of mosaic, it is important to get the shape of each tile. Another possibility is acquiring a digital image of the mosaic and applying an image segmentation algorithm to automatically detect the tiles thus building a "digital model" of the mosaic.

Dispose of a digital model of the mosaic could be a great advantage both in the traditional fields referring to the humanities and in the new technologies applied to the study, the preservation and the valorization of the worldwide cultural heritage. Indeed, a "digital model" could be useful for archaeologists, scholars and restorers who are interested in studying, comparing and preserving the mosaics. Moreover, for restorers a digital model of a mosaic could become an essential professional tool. Coming to the new technologies applied to the cultural heritage, the virtual reproduction of a mosaic is the first step to build the 3D model. This could be used to rebuild the place where the mosaic or the mosaics was/were originally set, helping in visualizing it for preservation and valorization use.

To the best of our knowledge, a single mosaic-oriented segmentation algorithm has been proposed in the literature (Benyoussef and Derrode, 2008) that is based on the well-known watershed algorithm (Vincent and Soille, 1991) and some mosaic-specific preprocessing. Unfortunately, its performance has been assessed subjectively only. On the other hand a number of recent general purpose segmentation algorithms exist (Moore et al., 2008; Levinshtein et al., 2009; Liu et al., 2011; Achanta et al., 2012), but still is not clear whether they are suitable for segmenting mosaics and how they comparatively perform. The purpose of the present paper is twofold: on the one hand, we suggest that a simple error metric taking into account precision, recall and the error on the number of tiles is appropriate for the objective evaluation of a mosaic segmentation. We further introduce some mosaic-specific hardness estimators (namely some indexes of how difficult is the task of segmenting a par-

Fenu G., Jain N., Medvet E., Pellegrino F. and Namer M..
On the Assessment of Segmentation Methods for Images of Mosaics.
DOI: 10.5220/0005310101300137
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 130-137
ISBN: 978-989-758-091-8

ticular mosaic image). On the other hand, we compare the performance of a number of popular segmentation algorithms based on a data-set that we collected and manually annotated thus providing a ground-truth that could be employed to assess performance of new mosaic segmentation algorithms.

The history of image segmentation evaluation dates back to the late seventies (Yasnoff et al., 1977). From then on, a plethora of evaluation methods, of both supervised and unsupervised nature have been proposed, see (Zhang et al., 2008) for a relatively recent survey. Here we employ a supervised (i.e., based on manually segmented images) evaluation method similar to the precision and recall method applied by (Neubert and Protzel, 2012), but with reference to the whole segment area instead of its boundary. Moreover, since the number of segments is an objective datum of the problem, it is reasonable to take into account the difference of the number of detected and ground-truth segments.

A survey, even partial, of segmentation methods is far beyond the purpose of this introduction. Here we only stress that a single segmentation method (Benyoussef and Derrode, 2008) specifically designed for mosaics has been proposed in the literature. The results reported in Section 3.1 have been obtained by using the original code kindly provided by the authors. As far as the other image segmentation algorithms employed in this paper are concerned, they all belong to the class of superpixel segmentation methods (Ren and Malik, 2003) meaning that the segmentation criterion is perceptual uniformity. In particular, we employed the Superpixel Lattice (SL) (Moore et al., 2008), the TurboPixels (TP) (Levinshtein et al., 2009), the Entropy Rate Superpixel Segmentation (ERSS) (Liu et al., 2011) and the SLIC Superpixels (SLIC) (Achanta et al., 2012). The segmentation methods will be briefly described in Section 3.1.

The reminder of the paper is organized as follows: in Section 2 the performance metrics and the hardness estimators are introduced. In Section 3.1 the segmentation methods, the dataset and the experimental results are reported. Finally, the conclusions are drawn in Section 4.

2 OBJECTIVE EVALUATION METRICS

Image segmentation is the process of partitioning a digital image into segments, representing possibly meaningful regions of the image. For the purpose of the present paper, a *segmentation method* $\mathcal{S}$ is a function such that $\mathcal{S}(p)$ is the label assigned to pixel $p \in I$, for a given image I. We allow a special case when the method deems the pixel not to belong to any meaningful region: we denote this case with $\mathcal{S}(p) = \varnothing$. Indeed, we believe that a segmentation method which is tailored to mosaic images should assign a $\varnothing$ label to the pixels corresponding to the filler, if any. A *region* $R \subseteq I$ is the largest set of pixels of a segmented image for which the label is the same and different from $\varnothing$. The *segmentation* $\mathcal{R}_I$ of the image I is the set of all regions.

A *ground truth* $\mathcal{T}_I$ is a set of non-overlapping pixel subsets of an image I; $\mathcal{T}_I$ represents the desired segmentation of a mosaic image. A *tile* $T \in \mathcal{T}_I$, with $T \subseteq I$, contains all and only the pixels corresponding to the same tessella on the mosaic depicted in I.

2.1 Performance Metrics

Given a segmentation $\mathcal{R}_I$ performed by a method $\mathcal{S}$ on an image I and the desired segmentation $\mathcal{T}_I$, we aim at determining quantitatively how close $\mathcal{R}_I$ is to $\mathcal{T}_I$. To this end, we define three metrics: average tile precision $\mathrm{Prec}(\mathcal{R}_I, \mathcal{T}_I)$, average tile recall $\mathrm{Rec}(\mathcal{R}_I, \mathcal{T}_I)$, tile count error $\mathrm{Count}(\mathcal{R}_I, \mathcal{T}_I)$. More in details:

$$\mathrm{Prec}(\mathcal{R}_I, \mathcal{T}_I) = \frac{1}{|\mathcal{T}_I|} \sum_{T \in \mathcal{T}_I} \frac{\max_{R \in \mathcal{R}_I} |R \cap T|}{|R|}$$

$$\mathrm{Rec}(\mathcal{R}_I, \mathcal{T}_I) = \frac{1}{|\mathcal{T}_I|} \sum_{T \in \mathcal{T}_I} \frac{\max_{R \in \mathcal{R}_I} |R \cap T|}{|T|}$$

$$\mathrm{Count}(\mathcal{R}_I, \mathcal{T}_I) = \frac{\mathrm{abs}(|\mathcal{T}_I| - |\mathcal{R}_I|)}{|\mathcal{T}_I|}$$

We chose to use precision and recall because the problem of segmenting mosaic images can be seen, to some degree, as an information retrieval problem: that is, given a mosaic tessella depicted on the image, how good is the segmentation method in finding all (recall) and only (precision) the relevant pixels, i.e., those which correspond to that tessella? In other words, precision and recall assess the method ability to correctly cover a tile with a region.

However, precision and recall alone cannot completely capture the segmentation method ability to meet the objective described in Section 1. For example, consider the case in which each tile in $\mathcal{T}_I$ is perfectly coincident to a region in $\mathcal{R}_I$, but there are several spurious regions in $\mathcal{R}_I$ which do not cover any tile—i.e., they cover the filler. Clearly, these regions will negatively affect the result of simple analysis of the mosaic performed on the segmented image, such as counting the number of tessellas or determining the amount of filler. Hence, in order to tackle this limitation, we defined the tile count error.

Finally, we also use the F-measure index, which is the harmonic mean of precision and recall:

$$\mathrm{Fm}(\mathcal{R}_I, \mathcal{T}_I) = 2\frac{\mathrm{Prec}(\mathcal{R}_I, \mathcal{T}_I)\mathrm{Rec}(\mathcal{R}_I, \mathcal{T}_I)}{\mathrm{Prec}(\mathcal{R}_I, \mathcal{T}_I) + \mathrm{Rec}(\mathcal{R}_I, \mathcal{T}_I)}$$

2.2 Hardness Estimators

We aim at quantifying those style-related respects characterizing a mosaic image which likely affect how difficult is the task of segmenting a particular mosaic image, i.e., at defining some *hardness estimators*. We considered four style features—tiles size variability, amount of filler, color dissimilarity within the tiles, color dissimilarity near to the edge of the tiles—and the corresponding indexes, which are defined in details below. We think that, in general, i.e., without making any assumption about the segmentation method which is being used, the lower the index, the easier the image segmentation—except for the last, for which the greater the index, the easier the segmentation.

We chose to consider the amount of filler as an hardness estimator because most of existing image segmentation methods usually do not allow for an image portion to not be covered by any region: however, when operating on mosaic images, this should be the ideal behavior for those image portions which corresponding to the filler. Hence, we think that the amount of filler may significantly impact on segmentation.

Of course, other image features may impact on segmentation hardness: e.g., sharpness, geometric distortion, resolution, etc. In this work, we focus on mosaic-related characteristics.

Each index can be computed starting from a segmentation of the image: in the following notation, we assume that the indexes are computed using the ground truth $\mathcal{T}$. We are not concerned here in providing some proxy for these indexes which can be computed without a segmentation.

Tile Size Variability. We define the index $\rho_{|T|} = \frac{\sigma_{|T|}}{\mu_{|T|}}$, where $\mu_{|T|}$ and $\sigma_{|T|}$ are the mean and standard deviations of the tiles size $|T|$, respectively, with $T \in \mathcal{T}$. $\rho_{|T|}$ captures the variability of tile size.

Filler Amount. We define the index $\rho_{\varnothing} = 1 - \frac{\sum_{T\in\mathcal{T}}|T|}{|I|}$, i.e., the relative size of the portion of the image which is not covered by any tile.

In-tile Color Dissimilarity. We define the index μ_{I_T} as the mean of the in-tile color dissimilarity $I_T = \sigma_{L^*} + \sigma_{a^*} + \sigma_{b^*}$, where σ_{L^*}, σ_{a^*} and σ_{b^*} are the standard deviation of the L^*, a^* and b^* channel values of the pixels within the tile T (in the CIE-Lab color space). I_T captures the dissimilarity of the color within a tile T and could be affected (i.e., have a large value) by deterioration, tile surface unevenness causing shadows, and so on.

Out-tile Color Dissimilarity. We define the index μ_{O_T} as the mean of the out-tile color dissimilarity O_T, which is computed, for each tile T, as follows. Let $\bar{T}$ be a supertile of T which has the same barycenter of T and is scaled by a factor $\beta > 1$. We set:

$$O_T = \left(\left(\mu^T_{L^*} - \mu^{\bar{T}\setminus T}_{L^*}\right)^2 + \left(\mu^T_{a^*} - \mu^{\bar{T}\setminus T}_{a^*}\right)^2 + \right.$$
$$\left. + \left(\mu^T_{b^*} - \mu^{\bar{T}\setminus T}_{b^*}\right)^2 \right)^{\frac{1}{2}}$$

where $\mu^T_{L^*}$ and $\mu^{\bar{T}\setminus T}_{L^*}$ are the mean values of the L^* channel values in the pixel of T and $\bar{T} \setminus T$, respectively, being $\bar{T} \setminus T$ the portion of the supertile $\bar{T}$ not covered by T (the same for a^* and b^*). O_T captures the dissimilarity of the mean color of the tile T from the mean color of the area adjacent to T. We set $\beta = 1.2$.

3 EXPERIMENTAL COMPARISON

3.1 Segmentation Methods

We considered 5 segmentation methods. One of them (Benyoussef and Derrode, 2008) is tailored specifically to mosaics and will be referred to as Tessella-Oriented Segmentation (TOS). The remaining 4 methods are of superpixel type, namely they aim to produce segments that are perceptually uniform. Due to the nature of mosaics, this kind of segmentation is desirable. The methods have been selected among:

- recent (posterior to 2008);
- fast (meaning that, with reasonable parameters and reasonable image size, the processing time is no more than a few minutes—for instance, the processing time[1] of SLIC on a 647 kpixel image is 83 s);
- whose code is publicly available for free.

[1]Mean execution time of the SLIC segmentation algorithm, using Matlab R2012a on MacBook Pro, with 8 GB RAM and CPU Intel Core 2 Duo, 3.06 GHz.

Based on the above requirements we selected the Superpixel Lattice (SL) (Moore et al., 2008), the TurboPixels (TP) (Levinshtein et al., 2009), the Entropy Rate Superpixel Segmentation (ERSS) (Liu et al., 2011) and the SLIC Superpixels (SLIC) (Achanta et al., 2012).

Tessella-Oriented Segmentation (TOS). The main idea of the algorithm is modelling explicitly the mortar joint, i.e., exploiting the fact that the tesserae are surrounded by the filler. More precisely, for each pixel, some directional gray-level profiles are calculated and compared to a parametrized model. In this way a "criterion image" to be processed by the watershed transform is obtained. The algorithm has a single parameter (α) corresponding to half of the average tile size.

Superpixel Lattice (SL). The algorithm relies on subsequent bipartitions of the image. The bipartitions are, alternatively, horizontal and vertical and such that the topology of a regular lattice is preserved. The optimal partition paths are chosen in order to maximize the cross-part difference, hence are likely to pass through actual image boundaries. The main parameter is the *resolution*, corresponding to the total number of superpixels, organized in a $l_w \times l_h$ grid.

TurboPixels (TP). TP is a geometric flow-based algorithm, where K circular seeds are distributed uniformly over the image and their boundaries evolve in time according to a diffusion-like equation. The evolution is governed by conflicting terms, called reaction-diffusion and "doublet", whose weights α and β respectively, have been assigned the values suggested by the authors. Thus, the algorithm depends on the parameter K only.

Entropy Rate Superpixel Segmentation (ERSS). In the ERSS algorithm, the whole image is represented by a graph, whose nodes correspond to the pixels, while the arc weights between adjacent nodes denote the pairwise similarities. Then, the segmentation problem is formulated as a graph partition problem, where K subgraphs (corresponding to K superpixels) are found in order to maximize a proper objective function. The objective function comprises two terms: the entropy rate of a random walk on a graph and a balancing term. The entropy rate favors formation homogeneous clusters, while the balancing function enforces clusters of similar size. Besides K the algorithm relies on two parameters: the weight λ' of the balancing term w.r.t. the entropy rate term and the width σ of the Gaussian kernel used for computing pixel similarity. According to the authors, $\lambda' = 0.5$ yields a good compromise between the opposite requirements, while the algorithm is little sensitive to σ for a wide range of values. Hence the most significant parameter is K.

SLIC Superpixels (SLIC). SLIC stands for Simple Linear Iterative Clustering and is a simple adaptation of the well-known k-means clustering algorithm. The distance measure employed combines color and spatial proximity. In order to speed-up calculations, the distances are computed in a neighborhood proportional to the superpixel size. The algorithm has two parameters, precisely the desired number of approximately equally sized superpixels and the weighting factor between colour and spatial proximity for the calculation of the distance measure. Based on extensive experimentation, the main parameter has been identified as the weighting factor w. Indeed, we noted that different weighting factor values may result in segmentations with very different number of superpixels, even if using the same initial guess for the desired number of superpixels.

3.2 Mosaics

We considered 5 mosaic images, depicting 5 mosaics which differ in age and style (see Figure 1):

Church. This is a portion of a mosaic forming the floor of the Basilica di Santa Maria Assunta, the principal church in the town of Aquileia (Udine, Italy). The original church dated back to the fourth century; the current basilica was built in the eleventh century and rebuilt again in the thirteenth century. The image has been acquired with a camera not perfectly orthogonal to the mosaic plane.

Museum. This is a portion of a mosaic which is displayed at the Early Christian Museum (Museo paleocristiano di Monastero) in Aquileia. The museum houses several mosaic fragments from different excavations of Aquileia.

Bird. This is a contemporary mosaic built by an Italian artist[2] inspired by mosaics of the aforementioned Basilica.

Flower. This is a small contemporary mosaic, built by an Italian amateur as an essay for a course of ancient mosaic technique.

[2]http://latenagliaimpazzita.wordpress.com/2014/03/04/aquileia-eia-eia-uh

Table 1: Salient information of the 5 considered mosaic images.

Mosaic image	Size	$\lvert\mathcal{T}_I\rvert$	$\lvert I\rvert$	$\mu_{\lvert T\rvert}$	$\rho_{\lvert T\rvert}$ [%]	$\rho_{\varnothing}$ [%]	μ_{I_T}	μ_{O_T}
Church	793×1031	855	0.82×10^6	543.5	59.5	43.2	39.2	806.9
Museum	1951×2386	135	4.66×10^6	21 451.8	40.1	37.8	13.8	201.0
Bird	942×704	717	0.66×10^6	530.0	58.0	42.7	54.3	618.2
Flower	1097×1872	546	2.05×10^6	2646.0	41.8	29.6	42.6	315.5
University	438×340	134	0.15×10^6	666.1	70.3	40.1	35.1	962.3

University. This is a portion of a mosaic forming the floor in the main building of our campus, which dates back to the 1938. This mosaic is, in several points, quite deteriorated due to pounding.

Table 1 shows the values of the hardness estimators indexes (leftmost 4 columns) for the 5 mosaic images, along with width and height of the image, the number $|\mathcal{T}_I|$ of tiles in the corresponding ground truth, the size of the image $|I|$, and the average size of a tile $\mu_{|T|}$.

3.3 Results and Discussion

We applied each method to each mosaic image, with several values for the main method parameter (totaling 603 method applications) and obtained the corresponding segmentations: with respect to the main parameter value, we sampled more densely those intervals in which the performance appeared to be better. We then computed the performance metrics presented in Section 2.

With the aim to harmonize the results and to easily compare the performances of different algorithm, we chose to scale the main parameter of each method (as briefly identified and described above) by the effective number of tiles in each mosaic, when appropriate. Indeed, the parameters of TP, ERSS and SL methods are the desired numbers of superpixels, hence scaling the parameters allows for performance comparison of every method applied on the full set of available mosaics. Instead, considering the other two algorithms (TOS and SLIC), the parameters are related to different mosaic information and the scaling by the effective tiles number would be useless. More in detail, let p be the main parameter: we set $p := \alpha$ for TOS, $p := \frac{l_w l_h}{|\mathcal{T}_I|}$ for SL, $p := \frac{K}{|\mathcal{T}_I|}$ for TP and ERSS, and $p := w$ for SLIC. For SL, we experimented, for each mosaic image, with l_w, l_h values consistent with the width/height ratio of the image.

The results are shown graphically in Figure 2 and are summarized in Table 2. Figure 2 consists of 20 charts, one row per method: the first column (left) shows the F-measure vs. the main parameter, the second column (center) shows the count error vs. the main parameter, and the third column (right) shows F-measure vs. count error. Charts of the first two columns show one line per mosaic image. Table 2 shows the performance metrics obtained by the methods on the mosaic images, using only the main parameter value which led to the best F-measure.

Table 2: Performance metrics of the segmentation methods obtained on the mosaic images (MI) using the main parameter value (third column) which led to the best F-measure. For each mosaic image, the best value for the count error and the best value for the F-measure are highlighted.

MI	Method	p	Count	Prec	Rec	Fm
Church	TOS	4.00	0.54	56.4	71.9	63.2
	SL	1.08	**0.08**	35.3	71.1	47.2
	TP	2.39	1.54	63.5	63.8	**63.7**
	ERSS	2.83	1.83	51.2	39.3	44.4
	SLIC	12.83	0.25	47.5	83.8	60.6
Museum	TOS	6.00	0.14	64.4	87.3	74.1
	SL	1.07	**0.07**	52.3	91.7	66.6
	TP	1.73	0.84	60.8	60.3	60.6
	ERSS	2.16	1.18	48.0	39.5	43.3
	SLIC	5.00	7.03	73.9	80.6	**77.1**
Bird	TOS	4.00	**0.03**	52.8	81.7	64.2
	SL	1.31	0.31	40.4	68.6	50.9
	TP	2.14	1.28	62.9	68.4	**65.5**
	ERSS	2.94	1.94	55.3	39.9	46.3
	SLIC	12.00	0.32	50.6	84.8	63.4
Flower	TOS	4.00	0.06	49.4	67.8	57.2
	SL	0.99	**0.01**	37.1	63.1	46.7
	TP	1.72	0.83	67.4	63.5	65.4
	ERSS	2.75	1.75	60.2	35.3	44.5
	SLIC	6.00	1.74	63.4	70.4	**66.8**
University	TOS	4.00	0.90	63.2	78.5	**70.1**
	SL	1.04	**0.04**	44.0	88.6	58.8
	TP	1.90	1.78	67.0	68.9	67.9
	ERSS	2.99	1.99	52.6	40.9	46.0
	SLIC	8.00	0.33	54.5	87.9	67.3

It can be seen that the best (largest) values in terms of F-measure are obtained with TOS and SLIC methods. In particular, these methods achieve 74.1% and 77.1% F-measure for the Museum image—best values for ERSS, TP and SL are respectively 46.3%, 67.9% and 66.6% on the Bird, University and Museum images. This finding is confirmed by Table 2, which highlights, for each mosaic, the best values obtained for the count error and the F-measure: concerning the latter, SLIC performs best on 2 images, TP on 2 and TOS on the last image; TOS is the second best

(a) Church (b) Museum (c) Bird (d) Flower (e) University

Figure 1: The mosaics images considered in our evaluation.

performing method in 3 images.

SL and TOS are the best performing methods in terms of the count error—note also the range of the y-axis of the corresponding charts. We recall that, for the former, the main parameter determines exactly which will be the number of regions in the segmentation: hence, setting $p = \frac{l_w l_h}{|\mathcal{T}_I|} = 1$ leads to a count error equal to 0. However, in a real scenario $|\mathcal{T}_I|$ could not be available, or could be estimated only roughly. On the other hand, for the TOS method, the main parameter does not depend on $|\mathcal{T}_I|$; morevoer the best values for the F-measure and count error metrics are obtained for (approximately) the same values for the main parameter.

This finding is further corroborated by the Fm vs. Count chart (third column of Figure 2): for the TOS method, the points with the largest F-measure and those with the smallest count error are quite close. Instead, for the other methods (in particular for the SLIC method), this chart shows a clear trade-off between F-measure and count error: in other words, the main parameter value which allows to obtain a good segmentation in terms of F-measure, also leads to a remarkably bad result in terms of count error (50% more regions found than the actual tiles). We think that this limitation is caused by the fact that methods other than TOS do not take into account the presence of the filler. Hence, they tend to introduce regions which cover the filler but negatively affect the count error.

Concerning the sensitivity of the methods w.r.t. their main parameters, results show that the count error is affected more distinctly than the F-measure. We remark that for 3 on 5 methods (SL, TP and ERSS) the main parameter represents the desired number of regions in the segmentation, i.e., the number of tiles: a reliable estimate for this figure could not be available in a real world scenario. For the other two methods (TOS and SLIC), performances apper to be little sensitive to reasonable variations in the main parameter.

Another interesting result concerns how the hardness estimators (see Section 2.2) impact on the method performances. The two best performing methods (TOS and SLIC) obtain the best F-measure values on the Museum and University images. Indeed, the Museaum image has the smallest value for the μ_{I_T} index and the University image has the largest value for the μ_{O_T} index (see Table 1). Despite the number of mosaic images which have been analyzed is not sufficient for drawing statistically significant conclusions, this finding appears to support the hypothesis that μ_{I_T} and μ_{O_T} can be used as good predictors of the performance of a mosaic image segmentation method. From another point of view, if one would aim at improving the performance of a segmentation method, he/she could augment the method with an image preprocessing step which decreases the value of μ_{I_T} and increases the value of μ_{O_T}. Recall, however, that the actual values for μ_{I_T} and μ_{O_T} cannot be computed without knowing the ground truth of the mosaic image to be analyzed.

Finally, it can be noted that ERSS and TP methods also perform relatively well on University image, whereas they perform poorly on the Museum image. We think that this can be explained by the fact that the latter image has a higher resolution and these methods are probably not robust w.r.t. the input image resolution.

4 CONCLUDING REMARKS AND FUTURE WORK

We considered the problem of the automatic segmentation of mosaic images. We proposed a set of simple performance metrics and some mosaic-specific hardness estimators. We assessed both on a set of recent segmentation methods which we applied to 5 manually annotated mosaic images. Our experimental analysis shows that: (i) methods which are not specifically tailored to the segmentation of mosaic images are hampered by not taking into account the filler, which introduces a trade-off between F-measure and count error; (ii) in-tile color dissimilarity and out-tile

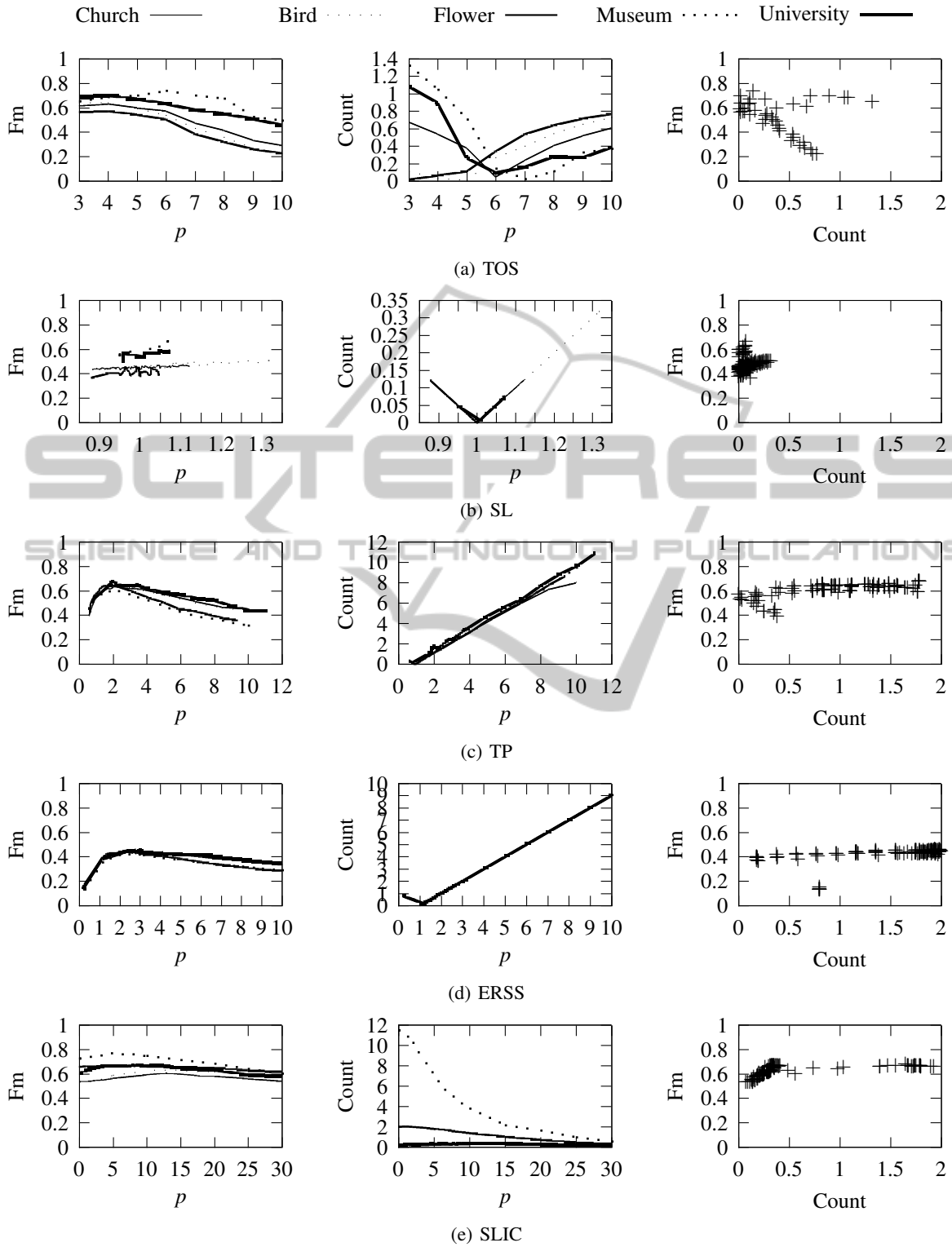

Figure 2: F-measure (left column), count error (central column) for the 5 methods (one row per method) on the 5 mosaic images (one line per image) vs. the methods main parameter (see text). The third column shows F-measure vs. count error: top-left corner is the ideal performance.

color dissimilarity appear to be good estimators for the hardness of the segmentation of a mosaic image; and (iii) the performance, in terms of F-measure, of the only method which is specifically tailored to mosaic images appear not significantly better than those of other, general-purpose segmentation methods.

ACKNOWLEDGEMENTS

This work has been funded by FRA 2012, Finanziamento d'Ateneo per progetti di ricerca scientifica, of University of Trieste. The authors would like to thank the Regional Archaeological Service (Soprintendenza per i Beni Archeologici del Friuli Venezia Giulia) for assistance during the photographic collection of mosaics in Aquileia. The authors would like to thank also Prof. Stéphane Derrode of Ecole Centrale Lyon, LIRIS, for providing the original code of (Benyoussef and Derrode, 2008).

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282.

Benyoussef, L. and Derrode, S. (2008). Tessella-oriented segmentation and guidelines estimation of ancient mosaic images. *Journal of Electronic Imaging*, 17(4):043014–043014.

Benyoussef, L. and Derrode, S. (2011). Analysis of ancient mosaic images for dedicated applications. In Stanco, F., Battiato, S., and Gallo, G., editors, *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. CRC Press.

Levinshtein, A., Stere, A., Kutulakos, K. N., Fleet, D. J., Dickinson, S. J., and Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297.

Liu, M.-Y., Tuzel, O., Ramalingam, S., and Chellappa, R. (2011). Entropy rate superpixel segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2097–2104. IEEE.

Moore, A. P., Prince, S., Warrell, J., Mohammed, U., and Jones, G. (2008). Superpixel lattices. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Neubert, P. and Protzel, P. (2012). Superpixel benchmark and comparison. In *Proc. Forum Bildverarbeitung*.

Ren, X. and Malik, J. (2003). Learning a classification model for segmentation. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 10–17. IEEE.

Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598.

Yasnoff, W. A., Mui, J. K., and Bacus, J. W. (1977). Error measures for scene segmentation. *Pattern Recognition*, 9(4):217–231.

Zhang, H., Fritts, J. E., and Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding*, 110(2):260–280.