Scene Text Extraction using K-means Clustering in HSI Color Space: Influence of Color Distance Measure

MATKO SARIC, MAJA STELLA, PETAR SOLIC
Faculty of electrical engineering, mechanical engineering and naval arhitecture
University of Split
R. Boskovica bb, Split
Croatia
msaric@fesb.hr, mstella@fesb.hr, psolic@fesb.hr

Abstract: Text extraction in scene text images is necessary step that enables higher recognition performance. Challenges in this task arise from wide set of degradations present in natural scene images like complex backgrounds, uneven illumination, viewing angle, etc. In this paper we investigate influence of color distance measure on scene text extraction performance of K-means algorithm in HSI color space. For comparison purposes we tested following clustering distances: hue distance, saturation distance, intensity distance, chromatic distance and cylindrical distance. Obtained results are analyzed with respect to their complementarity in order to show potential for performance improvement

Key-Words: scene text extraction, K-means, cylindrical distance, intensity distance

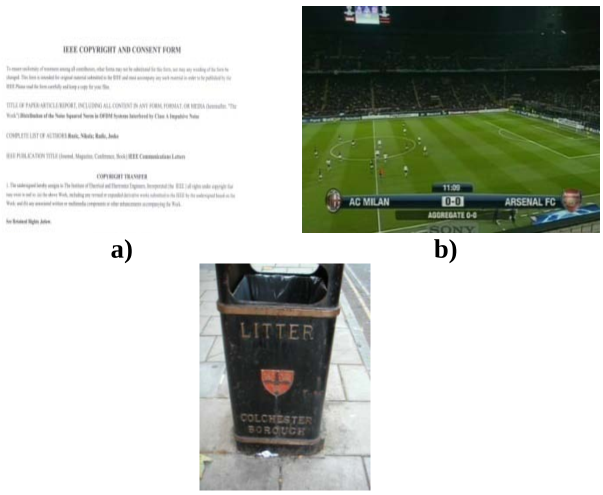
I. Introduction

Extraction of textual information from images and video enables range of different applications such as document analysis [1], automatic license plate recognition [2] - [3], sign detection and translation [4], road signs recognition [5], content based image indexing [6], etc. In literature text is divided on next categories: text in documents, caption text and scene text [7]. Text in documents (Fig. 1a) has properties (high contrast, uniform font and background etc.) that enable easier character extraction and recognition. Caption text (Fig. 1b) refers to characters artificially added to image or video frame. Typical examples are subtitles or match results in sports video. Scene text (Fig. 1c) is integral part of recorded image or video frame, that is text found in everyday environment (for example label on the door or text on traffic signs). This type of text is characterized by variability of background, shape, font, etc.

Digital cameras and camera equipped smartphones give users opportunity to take photo or record video almost wherever and whenever they want. This also means that scene text present in our environment can be easily captured. Despite these advantages, problems arising from usage of these devices refer to sensor noise, viewing angle, blur, variable illumination etc. These conditions, in combination with fact that scene text doesn't have constraints like text in documents, make its extraction a challenging task.

In [7] text information extraction procedure is divided on 5 steps: detection, localization, tracking, extraction and enhancement, and recognition (OCR). In extraction step characters are segmented from background, that is, text pixels are separated from background pixels. Artifacts resulting from poor character extraction (parts of background, missing character parts etc.) can significantly lower accuracy of recognition performed by OCR software. Importance of extraction step is especially emphasized in case of scene text. Complex backgrounds, geometrical deformation, uneven illumination and other degradations make character extraction a demanding task that determines success of recognition stage.

In [8] text extraction methods are divided in two categories: thresholding-based and grouping-based. Histogram thresholding [9], adaptive or local thresholding [10] and entropy based methods belong to first category. These methods have low computational requirements and successfully handle grayscale images or color channels separately, but they are not suitable in case of complex backgrounds and varying colors. Region-based, learning-based and clustering-based methods belong to second category. Region-based techniques includes methods like region-growing [11] - [12] and split-and-merge algorithm [13]. Their main advantage is inclusion of spatial information which is very important for character extraction, but parameter values have strong influence on



a) **b)** **c)**

Fig. 1a) text in document b) caption text c) scene text performance. In learning-based methods text extraction is performed using well–known classifiers (multi-layer perceptrons, self-organizing maps). Main problem is creation of representative training database in order to encompass variability of scene text examples.

Clustering-based methods are based on assumption that pixels have tendency to form groups in chosen color space. K-means is one of the most often used algorithms for text extraction because of its speed and efficiency. In [14] scene text is extracted using K-means clustering in RGB color space with Euclidean distance and angle distance where authors choose better result based on feedback from recognition results. Number of clusters is set to 3 representing characters, background and character edges. Garcia and Apostolidis [15] exploited 4-means for text segmentation in HSV color space. In [16] authors propose text extraction method based on K-means clustering with modified cylindrical distance in HSI color space. Comparison is also made with K-means using cylindrical distance in HSI color space and Euclidean distance in RGB color space. Wakahara and Kita [17] used K-means clustering for generation of multiple extraction results where best result is chosen using SVM classifier.

In this paper we investigate influence of color distance measure on scene text extraction performance of K-means algorithm in HSI color space. For comparison purposes we tested following clustering distances: hue distance, saturation distance, intensity distance, chromatic distance and cylindrical distance. First three measures refer to single component differences. Chromatic distance is calculated as pixel distance in HS plane and it reflects chroma difference. This measure is taken from the definition of cylindrical distance in HSI color space where overall pixel distance is calculated using chromatic and intensity distance. Cylindrical distance takes into account angular values and therefore it better corresponds to cylindrical nature of polar color spaces (HSV, HSI, HSL etc.) than Euclidean distance

Scene text extraction performance is evaluated for every single distance using total edit distance and correct recognition rate. We also analyze degree of complementarity between results obtained with different distances as possible direction to improve scene text extraction perrormance.

The rest of the paper is organized as follows. Section 2. shortly describes HSI color space and distance measures used in comparison. Section 3. discusses in more details K-means clustering. Results are presented in section 4 and conclusions are made in section 5.

II. Choice of color space and color distance measure

A. HSI color space

Unlike RGB color space, HSI (hue, saturation, intensity) color space corresponds to the human interpretation of colors. Hue is component determined by dominant wavelength of visible electromagnetic radiation. Saturation refers to the color purity where its lower value manifests as color closer to grey. Intensity refers to the perceived amount of light reflected or emitted from object. HSI color space is usually represented as double cone (Fig. 2). Hue is defined as angle difference to the red and it has value from the interval $\left[0^{\circ},360^{\circ}\right]$. Position on vertical axis represents intensity value, while distance from this axis is equal to saturation.

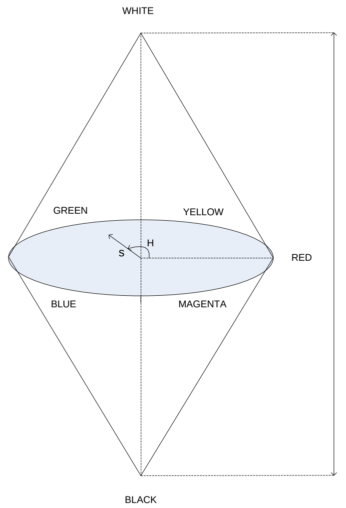


Fig. 2 HSI color space

Transformation from RGB to HIS color space is defined by:

$$H=\begin{cases}\alpha & if\ B\le G\\ 360^{\circ}-\alpha & if\ B>G\end{cases} \tag{1}$$

$$\alpha=\cos^{-1}\left\{\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}}\right\}$$

$$S=1-\frac{3}{(R+G+B)}\left[\min(R,G,B)\right] \tag{2}$$

$$I=\frac{1}{3}(R+G+B) \tag{3}$$

Besides fact that HSI color space corresponds to the way how human perceives colors, it also enables separation of chromatic (hue and saturation) and achromatic (intensity) information. Disadvantage is its perceptual nonuniformity meaning that Euclidean distance between two colors doesn’t correspond to perceived color difference. It is also important to mention following properties of hue and saturation components [18]:

- hue is meaningless when intensity is very low,
- hue is unstable when saturation is very low,
- saturation is meaningless when intensity is very low.

Lowering of the intensity value increases standard deviation of hue and saturation, while at low saturation values hue has larger standard deviation.

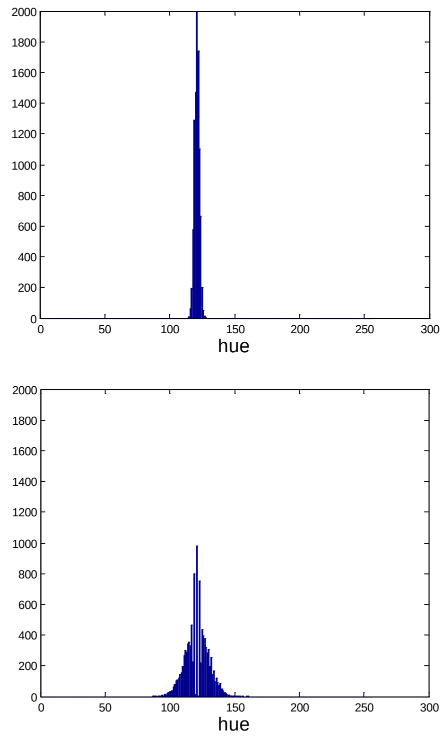


Fig. 3 Hue distribution for S=100 (top) and S=20 (down)

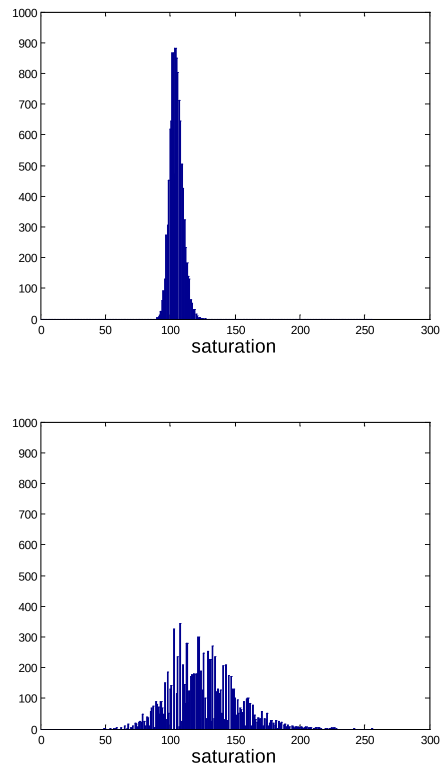


Fig. 4 Saturation distribution for I=100 (top) and I=20 (down)

This means that hue and saturation aren't relevant attributes in cases of low intensity and saturation levels. On Fig. 3 hue distribution in dependence on saturation values is shown. Hue histogram is computed from RGB sample (size 100x100) corrupted with additive white Gaussian noise where every component has the same standard deviation ($\sigma = 3$). The initial hue and intensity values of the sample are 120 and 100. Results of similar analysis for saturation component in HSI color space is shown in Fig. 4. where initial values of hue and saturation are 120 and 100.

It should be noted that hue and saturation are robust to highlights and shadowing and hence they have potential to correctly extract characters in presence of such degradations. This is also the reason why HSI color space is chosen for scene text extraction in this paper.

B. Color distance measures

Purpose of color distance measures is quantification of color difference. Choice of color distance has strong influence on text extraction performance. Representation of colors as 3D vectors enables usage of well-known distance measures for m-dimensional vectors. One solution is usage of Minkowski metric:

$$d_p(i,j) = c\left(\sum_{k=1}^{m} \xi_k \left|x_{ik} - x_{jk}\right|^p\right)^{\frac{1}{p}} \qquad (4)$$

where m is dimension of vector, x_{ik} and x_{jk} are k-th elements of $\vec{x}_i$ and $\vec{x}_j$. The nonnegative scaling parameter c represents discrimination power, while parameter ξ_k is weight of component k ($\sum_{k=1}^{m} \xi_k = 1$). Value of parameter p determines distance measures as city-block or Manhattan distance ($p = 1$), Euclidean distance ($p = 2$) and Chessboard distance ($p = \infty$).

These measures reflect difference between color vectors magnitudes, but it is also interesting to take into account angle distances. One such measure is derived from normalized inner product of color vectors:

$$\theta = \cos^{-1}\left(\frac{\vec{x}_i \vec{x}_j^{\,T}}{\left|\vec{x}_i\right|\left|\vec{x}_j\right|}\right) \qquad (5)$$

where $\vec{x}_i$ and $\vec{x}_j$ represent vectors and θ is angle between them. Approach combining these two kinds of distances was proposed in [14] where authors used Euclidean distance and angle distance between RGB colour vectors for scene text extraction task.

Although Euclidean distance is often used in RGB color space, it is not appropriate color difference measure for HSI color space because of its cylindrical nature. Cylindrical distance, introduced in [18], takes into account angle differences and hence in this paper it is chosen as one of the distance measures in K-means clustering. For two

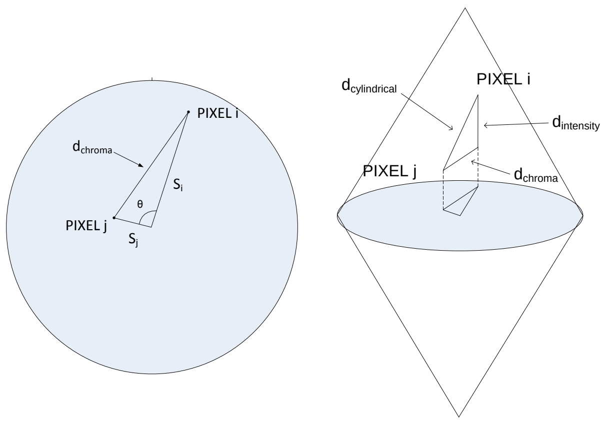


Fig. 5 Chromatic distance (a) and cylindrical distance (b)

pixels in HSI color space with values (H_i, S_i, I_i) and (H_j, S_j, I_j) cylindrical distance is defined as:

$$d_{cylindrical}(i,j) = \sqrt{d_{chroma}^2(i,j) + d_{intensity}^2(i,j)} \quad (6)$$

$$d_{chroma}(i,j) = \sqrt{S_i^2 + S_j^2 - 2S_iS_j\cos\theta} \quad (7)$$

$$d_{intensity}(i,j) = |I_i - I_j| \quad (8)$$

$$\theta = \begin{cases} \Delta & if\ \Delta < 180^\circ \\ 360^\circ - \Delta & otherwise \end{cases} \quad (9)$$

$$\Delta = |H_i - H_j| \quad (10)$$

where $H \in [0^\circ, 360^\circ]$, $S \in [0, 255]$, $I \in [0, 255]$, $d_{chromatic}$ refers to chromatic distance (Fig. 5 left) between two pixels, while $d_{intensity}$ refers to absolute value of intensity difference (intensity distance). Cylindrical distance $d_{cylindrical}$ (Fig. 5 right) is represented as hypotenuse of right-angled triangle.

Hue difference (hue distance) θ is defined with equations (9) and (10) that incorporate cyclic property of hue component. Because of of hue and saturation instability at low intensity and saturation levels it is necessary to perform distinction between chromatic pixels and achromatic pixels. For chromatic pixels, that have stable hue and saturation, cylindrical distance is calculated according to equations (6)-(10). For achromatic pixels intensity is only relevant component. This implies that chromatic distance is unreliable and cylindrical distance is reduced to intensity difference $d_{intensity}$ (8).

Chromatic and achromatic pixels are usually classified by thresholding of saturation and intensity component. Achromatic pixels are defined with following conditions:

$$\begin{aligned} &intensity > 90\% \text{ max } intensity\, value, \\ &intensity < 10\% \text{ max } intensity\, value, \\ &saturation < 10\% \text{ max } saturation\, value. \end{aligned} \quad (11)$$

For comparison purposes in this paper saturation distance is also used:

$$d_{saturation}(i,j) = |S_i - S_j| \quad (12)$$

Besides intensity difference, cylindrical distance also considers hue and saturation difference through chrominance information.

III. Text extraction using K-means clustering

According to [8], clustering algorithms are considered as very efficient methods for scene text extraction. One of the most popular clustering techniques is K-means. Its main advantages are easy implementation and low computational requirements. K-means tries to minimize sum of distances between points and cluster centers that is represented by:

$$\sum_{j=1}^{k} \sum_{i \in S_j} distance\left(x_i^{(j)}, c_j\right) \quad (13)$$

where *distance* is the chosen distance measure between point $x_i^{(j)}$ and the cluster centre c_j, S_j is set containing elements of cluster j and k is number of clusters. First step is to choose k points as initial cluster centers (centroids) c_j

Algorithm consists of following steps:

1. In set of N points, corresponding to image pixels, choose k points as initial cluster centers (centroids) c_j

2. Assign each point to nearest cluster S_j based on its distance from cluster center c_j.

3. For each cluster S_j, compute a mean μ_j of each cluster and set the mean as new cluster center $(c_j = \mu_j)$

4. Repeat the steps 2 and 3 until the centroids no longer move

Flowchart of this algorithm is shown in Fig. 6. It should be noted that one of the main drawbacks of this algorithm is the need to fix the number of clusters. Regarding text extraction task 2 clusters (one representing character and second background) seems as logical choice, although in [14] authors used 3 clusters where third one corresponds to character edges.

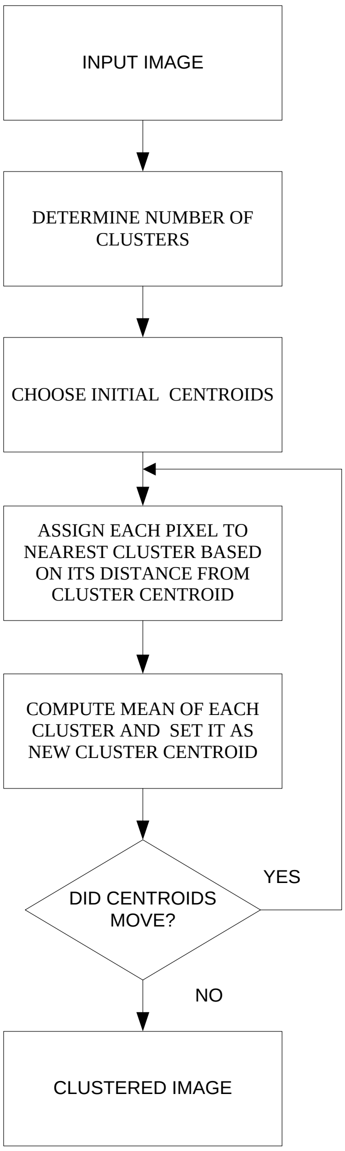


Fig. 6 Flowchart for k-means algorithm

In [19] author discussed a role to distance measure in K-means algorithm regarding text extraction. It is concluded that for RGB color space Euclidean distance gives best results, but angle distances can handle cases when Euclidean distance fails. Authors also tested performance of K-means algorithm with Euclidean distance measure in different color spaces where RGB yields best results. It should be mentioned that Euclidean distance is not appropriate measure for cylindrical color spaces (HSI, HSV, HLS etc.).

IV. Results

For evaluation purposes we used a test set from word recognition task in ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images [20]. This database contains 1189 scene text images covering a broad set of problems like non-uniform backgrounds, different layouts, low contrast, variable illumination, low resolution etc. MATLAB was used for implementation of text extraction based on K-means algorithm with different distances in HSI color space. The number of clusters is set to 2 (text and background). Binary images resulting from text extraction were processed with Google OCR engine Tesseract 3.02 in order to finally obtain recognized text.

As it is suggested in [20], the metrics for performance evaluation are edit (Levenshtein) distance and correct recognition rate. The first one is distance between ground truth string and string recognized using Tesseract, where deletions, substitutions and insertions have equal costs. Normalization is done by the number of characters in ground truth word. Second measure is percentage of correctly recognized words, that is, the number of words for which normalized edit distance is equal to zero.

Table 1 shows results of K-means clustering for hue distance, saturation distance, intensity distance, chromatic distance and cylindrical distance. First column shows results reported in form of total edit distance calculated by summing normalized edit distance for each ground truth word. Second column presents correct recognition rate.

Table I Scene text extraction results

Distance measure	Total Edit distance	Correct Recognition(%)
Hue distance	1306.2	10
Saturation distance	1153.5	14.1
Intensity distance	648.3	43.8
Chromatic distance	1083.2	17.4
Cylindrical distance	699.1	40.5

It can be seen that intensity distance outperforms other distances in both measures. This result reveals that intensity difference plays very important role in segmentation of characters from background. This observation is in accordance with conclusions presented in [19] and [16]: Euclidean distance, which mostly reflects changes in intensity component, performs best in scene text extraction task. In comparison with hue distance, which gives worst results, usage of saturation distance increases correct recognition rate and lowers edit distance. This can be explained by fact that saturation is more stable component: hue is unstable for low saturation and low intensity levels, while standard deviation of saturation increases only for low intensities. Chromatic distance includes hue and saturation differences and in comparison with results obtained with saturation distance, it gives higher value of correct recognition rate and lower edit distance. Cylindrical distance considers chroma and intensity difference, but this combination doesn't improve results in comparison with intensity distance only. Best result for this task on ICDAR 2011 database is reported in [21] where value of total edit distance is 639.15 and correct recognition rate is equal to 46.9%. Our results show that K-means clustering with intensity distance has slightly lower performance, but it is computationally less demanding.

Despite fact that hue, saturation and chromatic distances give obviously worse performance than intensity distance, it is interesting to investigate complementarity between these results: whether these measures, including chromatic information, are able to extract characters in cases when intensity fails? This assumption is inspired by fact that hue and saturation, unlike intensity, are robust to highlights and shadowing and therefore could enable character segmentation in presence of such degradations.

Analysis is performed to test complementarity between following distance pairs: intensity and chromatic distance, intensity and hue distance, intensity and saturation distance. For each pair it is analyzed in how many examples from test set one distance outperforms other in terms of edit distance and correct recognition rate. Table 2. shows that in 141 (11.9%) images from test set hue distance results with lower edit distance and correctly segments 33 (2.8%) words failed by intensity distance (Fig. 7). In table 3. it can be seen that saturation distance performs better than intensity difference (Fig. 8) in 125 (10.5%) cases according to edit distance and in 17 (1.4%) cases according to correct recognition rate. Finally, compared to intensity distance, chromatic distance (table 4.) gives lower edit distance (Fig. 9) for 162 (13.6%) examples and correctly extracts 41 (3.4%) words. Approach where examples that are not correctly extracted with intensity distance, would be segmented with chromatic distance could improve correct recognition rate for approximately 3% (from 43.8% to 47.3%).

This analysis confirms complementarity between tested distances that can be exploited to further improve performance by combination of results. From results presented in tables II.-IV. it can be concluded that chromatic distance shows highest level of complementarity with intensity distance. Despite overall efficiency of intensity distance, in cases when it fails chromatic distance is potential solution. This happens in images with certain kind of degradations like uneven illumination or shadows where robustness of hue and saturation plays important role.

Table II Complementarity between hue distance and intensity distance: *e_d_hue* is edit distance obtained for hue distance, *e_d_intensity* is edit distance obtained for intensity distance, *extracted_hue* is set of word correctly extracted with hue distance, *extracted_intensity* is set of words correctly extracted with intensity distance

	Number of words
e_d_intensity <e_d_hue	651 (54.8%)
e_d_intensity> e_d_hue	141 (11.9%)
$extracted\_intensity \cap \overline{extracted\_hue}$	554 (46.6%)
$\overline{extracted\_intensity} \cap extracted\_hue$	33 (2.8%)

Table III Complementarity between saturation distance and intensity distance: *e_d_ saturation* is edit distance obtained for saturation distance, *e_d_intensity* is edit distance obtained for intensity distance, *extracted_ saturation* is set of word correctly extracted with saturation distance, *extracted_intensity* is set of words correctly extracted with intensity distance

	Number of words
e_d_intensity <e_d_saturation	574 (48.3%)
e_d_intensity> e_d_ saturation	125 (10.5%)
$extracted\_intensity \cap \overline{extracted\_saturation}$	370 (31.1%)
$\overline{extracted\_intensity} \cap extracted\_saturation$	17 (1.4%)

Table IV Complementarity between chromatic distance and intensity distance: *e_d_intensity* is edit distance obtained for intensity distance, *e_d_chroma* is edit distance obtained for chromatic distance, *extracted_intensity* is set of word correctly extracted with intensity distance, *extracted_chroma* is set of words correctly extracted with chromatic distance

	Number of words
e_d_intensity<e_d_chroma	544 (45.8%)
e_d_intensity>e_d_chroma	162 (13.6%)
$extracted\_intensity \cap \overline{extracted\_chroma}$	355 (29.9%)
$\overline{extracted\_intensity} \cap extracted\_chroma$	41 (3.4%)

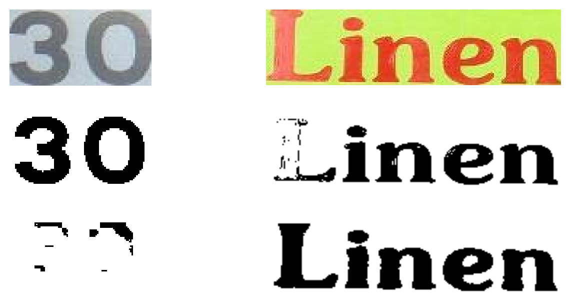


Fig. 7. Complementarity between intensity and hue distance: original images (top), result of K-means with intensity distance (middle) and result of K-means with hue distance (down)

Fig. 8. Complementarity between intensity and saturation distance: original images (top), result of K-means with intensity distance (middle) and result of K-means with saturation distance (down)

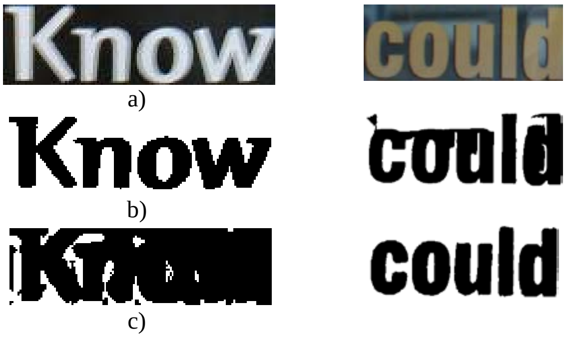


Fig. 9. Complementarity between intensity and chromatic distance: original images (top), result of K-means with intensity distance (middle) and result of K-means with chromatic distance (down)

V. Conclusion

In this paper we investigate how choice of color distance measure influences on scene text extraction performance of K-means algorithm in HSI color space. It is shown that intensity distance in K-means gives best results outperforming other tested distances. Obtained results confirm conclusions previously presented in literature [19] that intensity or lightness is the component that most efficiently segments characters from background in case of scene text images. We also analyzed complementarity between results obtained with tested distances: although intensity distance has best performance, hue, saturation and chromatic distance can successfully extract characters in some cases when intensity difference fails. This can be explained by fact that hue and saturation components are robust to highlights and shadows, while intensity is greatly affected by such degradations. In this way we show importance of chromatic information for scene text extraction. From presented results it can be concluded that fusion of results obtained with two or more distance measures in K-means algorithm has potential to improve scene text extraction performance.

References

[1] Y. Y. Tang, S. W. Lee, and C. Y. Suen, "Automatic document processing: a survey," *Pattern Recognition*, vol. 29, no. 12, p. 1931–1952.

[2] S. L. Chang, L. S. Chen, Y. C. Chung, and S. W. Chen, "Automatic license plate recognition," *IEEE Transactions on Intelligent Transport Systems*, vol. 5, no. 1, pp. 42-53, 2004.

[3] A. A. Safi, M. Azam, S. Kiani, and N. Daudpota, "Online vehicles license plate detection and recognition system using image processing techniques," in *Proceedings of the 5th WSEAS International Conference on Applied Computer Science*, Hanghzou, pp. 793-800.

[4] Y. Watanabe, K. Sono, K. Yokomizo, and Y. Okada, "Translation camera on mobile phone," in *Proceedings of International Conference on Multimedia and Expo*, 2003, pp. 177-180.

[5] A. V. Reina, R. J. L. Sastre, S. L. Arroyo, and P. G. Jiménez, "Adaptive traffic road sign panels text extraction," in *Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Robotics and Automation*, Madrid, 2006, pp. 295-300.

[6] M. Saric, H. Dujmic, V. Papic, N. Rozic, and J. Radic, "Player Number Recognition in Soccer Video using Internal Contours and Temporal Redundancy," in *Proceedings of the 10th WSEAS International Conference on Automation & Information (ICAI'09*, 2009, pp. 175-180.

[7] K. Jung, K. Kim, and A. Jain, "Text Information Extraction in Images and Video: A Survey," *Pattern Recognition*, vol. 37, no. 5, pp. 977-997, 2004.

[8] C. Mancas-Thillou and B. Gosselin, "Natural Scene Text Understanding," in *Vision Systems: Segmentation and Pattern Recognition*. Vienna, Austria: I-Tech Education and Publishing , 2007, ch. 16, pp. 307-332.

[9] S. Messelodi and C. M. Modena, "Automatic identification and skew estimation of text lines in real scene images," *Pattern Recognition*, vol. 32, no. 5, p. 791–810, 1999.

[10] B. Gatos, I. Pratikakis, K. Kepene, and S. J. Perantonis, "Text detection in indoor/outdoor scene images," in *Proc. First Workshop of Camera-based Document Analysis and Recognition*,, 2005, p. 127–132.

[11] R. Lienhart and W. Effelsberg, "Automatic text segmentation and text recognition for video indexing," University of Mannheim, Technical Report, 1998.

[12] H. Dujmic, M. Saric, and J. Radic, "Dujmić, Hrvoje; Šarić, Matko; Radić, Joško," in *Recent Researches in Neural Networks, Fuzzy Systems, Evolutionary Computing and Automation (Proceedings of 12th WSEAS conference on Automation & Information)* , 2011.

[13] D. Karatzas and A. Antonacopoulos, "Colour text segmentation in web images based on human perception," *Image and Vision Computing*, vol. 25, no. 5, pp. 564-577, 2007.

[14] C. Mancas-Thillou and B. Gosselin, "Color text extraction with

selective metric-based clustering," *Computer Vision and Image Understanding*, vol. 107, no. 1-2, pp. 97-107, 2007.

[15] C. Garcia and X. Apostolidis, "Text detection and segmentation in complex color images," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2000, p. 2326–2330.

[16] M. Saric, H. Dujmic, and M. Russo, "Scene Text Extraction in HSI Color Space using K-means Algorithm and Modified Cylindrical Distance," *Przegląd elektrotechniczny*, vol. 89, no. 5, 2013.

[17] T. Wakahara and K. Kita, "Binarization of Color Character Strings in Scene Images Using K-Means Clustering and Support Vector Machines ," in *2011 International Conference on Document Analysis and Recognition (ICDAR)* , 2011, pp. 274-278.

[18] D. C. Tseng and C. M. Chang, "Color segmentation using perceptual attributes," in *Proc. of the 11th Internat. Conf. on Pattern Recognition*, 1992, pp. 228-231.

[19] C. Mancas-Thillou, "Natural Scene Text Understanding," PhD Thesis, Faculté Polytechnique de Mons, 2006.

[20] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," in *Proc. 11th International Conference of Document Analysis and Recognition*, 2011, pp. 1491-1496.

[21] G. A and L. M. Bergasa, "A text reading algorithm for natural images," *Image and Vision Computing*, vol. 31, pp. 255-274, Mar. 2013.