Object Localization Using Color Histograms

Krzysztof Horecki[a0] & Dietrich Paulus[b] & Konrad Wojciechowski[a]

[a] Schlesische Technische Universität
Gliwice, Polen
krzyho@olimp.com.pl
http://www.olimp.com.pl/krzyho

[b] Lehrstuhl für Mustererkennung
Universität Erlangen–Nürnberg
paulus@cs.fau.de
http://www.mustererkennung.de

Abstract

In the contribution we present the comparison of different histogram distance measures and color spaces applied to object localization. In particular we examine the *Earth Mover's Distance* which has not yet been compared to other measures in the area of object localization. The evaluation is based on more than 80,000 experiments. Furthermore, we propose an extension to *Normalized RG* color space and an efficient scanning algorithm for localization.

1 Introduction

Localizing an object in a scene is a common task in computer vision. Color as a cue for solving this problem was presented by Swain and Ballard in [Swa91]. There it was shown that the *distributions of color information*, i.e. color histograms, can be used to solve the object localization problem.

In order to localize the object in the scene, we try to find that sub–image of the scene image, which has the smallest distance from the object image. The distances in sense of histograms are concerned.

We examined several histogram distance measures and color spaces in which we evaluated the histograms. We restricted our experiments to finding objects in the office scenery, whose sizes were known *a priori* from manual segmentation.[1]

This contribution is organized as follows; in Section 2 we introduce color histograms. In Section 3 we briefly describe the color spaces that were used in the experiments. Distance measures for histograms are discussed in Section 4. Section 5 describes how we applied histograms to object localization in real–world scenes. Section 6 describes the conducted experiments and presents the results. In Section 7 we summarize our contribution.

2 Color Histograms

There is no straightforward solution for the best color image feature representation. Adaptive histograms are said to describe the color information best. At the same time they need time–expensive clustering algorithms. [Rub98] used them for database image retrieval, i.e. an area well suited for off-line computations of histograms.

The requirements for the localization of objects differ from those of image retrieval. Several thousand histogram computations and their comparisons have to be performed on-line

[0] This work was carried out during a visit to the University Erlangen-Nürnberg, supported by the Socrates/Erasmus student program grant.

[1] In experiments published in [Ahl98], it was shown that the size can be estimated automatically from 3D information. [Vin98] proposed focused color intersection, where the objects can be localized without knowing their size in the scene.

for a single standard scene. That is why we took advantage of full histograms with uniform binning which can be quickly computed.

However, there are some drawbacks of this strategy. Uniform histograms do not accommodate the actual distribution of color information. E.g. for fine quantization many bins are empty or of no importance.

We assume that a close–up view of an object is given in the image $\boldsymbol{f} = [\boldsymbol{f}_{ij}]_{i=1\ldots M, j=1,\ldots N}$ where $\boldsymbol{f}_{ij}$ is a color pixel, i.e. $\boldsymbol{f}_{ij} = (r_{ij}, g_{ij}, b_{ij})^{\mathrm{T}}$. This object is to be found in the scene $\boldsymbol{f}'$. We compute histograms $\boldsymbol{S}$ for the scene and $\boldsymbol{T}$ for the object

$$\boldsymbol{S} = [S_l]_{l=1\ldots L} \quad \boldsymbol{T} = [T_l]_{l=1\ldots L} \tag{1}$$

where the number of bins L depends of the chosen quantization and the color space. The h function maps a color pixel to the index in the histogram, e.g. for a RGB histogram with $4 \times 4 \times 4$ bins ($L = 64$) and for color components in the range from 0 to 255 we might choose:

$$h : \begin{cases} \mathrm{I\!R}^3 & \rightarrow \quad \{1, \ldots L\} \\ \boldsymbol{f}_{ij} = (r_{ij}, g_{ij}, b_{ij})^{\mathrm{T}} & \rightarrow \quad \lceil r_{ij}/64 \rceil * 16 + \lceil g_{ij}/64 \rceil * 4 + \lceil b_{ij}/64 \rceil \end{cases} \tag{2}$$

The elements of the histogram $\boldsymbol{T}$ are defined as:

$$T_l = \frac{1}{NM} \left| \left\{ (i,j) | h(\boldsymbol{f}_{ij}) = l, i = 1 \ldots M, j = 1, \ldots N \right\} \right| \tag{3}$$

The elements of $\boldsymbol{S}$ are defined analogously.

3 Color Spaces

The histograms of color images can be computed in arbitrary color space. We decided to carry out our experiments using the following spaces:

RGB - the standard color space representing the red, green, blue components; $r, g, b \in [0, 1]$.

(Normalized) RG - eliminates intensity from the standard RGB model. Its components are defined as:

$$(r', g') = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b} \right) \tag{4}$$

HSI - the classical hue-saturation-intensity space. There is no unique conversion algorithm between RGB and HSI. We chose the one presented in [Fol92], because of its computational simplicity in comparison with algorithms presented in other sources.

YUV - color space family used in analog television transmission. Following the *ITU (International Telecomunication Union) R-601 Recommendation* , we used the YCbCr standard.[2]

CIElab - device-independent color space standard, which was derived from psychophysical experiments, describing all possible colors. Its vital property is that the geometric distance measured in it corresponds to the perceived differences between colors. We use the *ITU R-601 Recommendation* (with *D65 white illuminant*).

[2]Description can be found at `http://www.well.com/user/rld/vidpage/color_faq.html`

In our experiments we also use following chromaticity-only versions of the presented compound spaces:

H - consisting only of hue component of the HSI space,

UV - YUV model with luminance (Y) component removed,

AB - consisting only of CIElab chrominance components (a and b).

Apart from RGB, all other spaces do not cover the whole range of uniform binning space. We considered the normalized rg color space which covers only half of the area being binned. To make up for this problem, the binning strategy can be altered. The other solution is to transform the color space so that it covers the whole uniformly binned space. We found and propose the appropriate transformation for that space:

$$r'' = \begin{cases} r' + g' & : \text{ if } r' \geq b' \\ 2r' & : \text{ otherwise} \end{cases} \quad (5) \qquad g'' = \begin{cases} 2g' & : \text{ if } r' \geq b' \\ r' + g' & : \text{ otherwise} \end{cases} \quad (6)$$

The resulting space, after applying the described transformations, is presented in Fig. 2. We included the modified RG space now called *RG2* in the experiments.

4 Distance Measures

To be able to compare the histograms we need distance measures. In the following we present those which we chose to use in our experiments.

Histogram Intersection - a classical measure proposed in [Swa91]. It is computationally inexpensive generalization of geometric L_1 *Minkowski's distance* defined by:

$$\cap(\boldsymbol{S}, \boldsymbol{T}) = \sum_{l=1}^{L} \min\{S_l, T_l\} \tag{7}$$

Sum of Squared Differences (SSD) - defined by:

$$SSD(\boldsymbol{S}, \boldsymbol{T}) = \sum_{l=1}^{L} (T_l - S_l)^2 \tag{8}$$

Chi–square Test - statistical method. There are a few different versions of that measure in the literature [Puz98, Sch97, Pre88] We decided to use the following definition:

$$\chi^2(\boldsymbol{S}, \boldsymbol{T}) = \sum_{l=1}^{L} \frac{(S_l - T_l)^2}{T_l} \tag{9}$$

It was shown that using this definition, the best results were obtained with respect to noise, blur, and image plane rotation [Sch97].

Earth Mover's Distance (EMD) - Innovative measure presented in [Rub98] which represents the minimal transportation cost of one histogram to another.

$$\text{EMD}(\boldsymbol{S}, \boldsymbol{T}) = \min_{\boldsymbol{F}} \frac{\sum_{\mu=1}^{L} \sum_{\nu=1}^{L} F_{\mu\nu} D_{\mu\nu}}{\sum_{\mu=1}^{L} \sum_{\nu=1}^{L} F_{\mu\nu}} \tag{10}$$

EMD makes use of the ground distance D_{ij}, that is the distance between two histogram bins S_i and T_j; we chose the geometric distance between colors in particular color space for that task.

Computation of EMD involves the solution of the *transportation problem*, where a matrix $\boldsymbol{F} = [F_{\mu,\nu}]_{1,\ldots,L,1,\ldots,L}$ represents the flow from bin S_μ to T_ν. The flow cannot be negative ($0 \leq F_{\mu,\nu}$) and is further constrained by:

$$\sum_{\mu=1}^{L} F_{\mu\nu} = T_\nu \qquad \sum_{\nu=1}^{L} F_{\mu\nu} = S_\mu \qquad \sum_{\mu=1}^{L}\sum_{\nu=1}^{L} F_{\mu\nu} = 1 \quad . \tag{11}$$

It makes EMD the most computationally expensive among exploited measures. It can be implemented by means of linear programming methods, such as the simplex algorithm.[3]

5 Object Localization

In order to localize the object we performed an exhaustive search for the minimal distance between the object and the sub–image of the same size in the scene, where the distance between the histograms were concerned. This requires that the histogram of the estimated object size is computed at many locations in the scene image.

We exploited the *lower bounds* estimation algorithms to reduce the number of time-consuming distance measure computations. These algorithms are permitted to cancel histogram comparison when some threshold of difference is exceeded [Vin98, Rub98]. The lower bounds can be efficiently computed and they do not sacrifice accuracy of the results. We used *the Active Search* method [Vin98] for histogram intersection and *the distance between centroids of histograms* for EMD [Rub98].

Additionally we reduced the run–time for the local histograms computations at the hypothesized object location in the scene. Histograms of neighboring sub–images are highly correlated because of their relatively large common part. We reused previously computed histograms of the neighborhood by adding new pixels and subtracting those which no longer belonged to the sub–image.

For efficient histogram computation, we used a scan path that is shown in Fig. 1 which minimizes the number of required computations. This strategy resulted in a computational complexity which was independent of the histogram size.

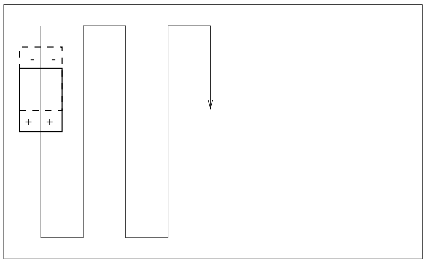

Fig. 1: Scan path.

In addition to the best match, four other possible object locations were estimated in the scene for each object. For that purpose, an interest map was computed with the size of the scene image. Each entry in this map represent the distance between the object histogram and the histogram of a sub–image around the given location in the scene image. Local maxima in this interest map represent possible object locations. Fig. 3 shows a scene in which the object 11 was localized. The interest map for a full search using the SSD distance measure is shown in Fig. 4. Fig. 5 shows the result of active search using histogram intersection.

[3]We used the Y. Rubner's code for EMD found at
`http://robotics.stanford.edu/ rubner/emd/default.htm`.

6 Experiments

We selected 17 objects as a data set for our experiments. The acquisition procedure consisted of the following steps:

- taking a picture of the object on a black background with some other objects in the scene,
- taking a picture of the same scene but with the object of interest removed,
- subtracting those images, then applying the median filter and thresholding it in order to obtain the mask,
- computing the bounding box of a mask of the object,
- finally, applying the AND operator to the mask and the first image mentioned to extract the object.

The result of this automatic procedure are images which are clamped to the mask boundary as shown in Fig. 6.

Next, we took images of 9 office scenes, each under four different illumination conditions. Each scene contained each of the object from the data set. To simulate the spectral illumination changes, we used daylight, bulb lamps and neon lamps. In one scene, we used a black cloth as a relatively homogeneous background.

We conducted localization for every color space, distance measure and histogram quantization combination. For Histogram Intersection, SSD and χ^2, the histograms with 6, 12, 16 and 32 quanta along each axis were used. In case of EMD we assumed an upper limit of a total of 100 bins, because of its high computational cost. It gave us 4 quanta for three-dimensional spaces (RGB, YUV, CIElab, HSI) and 4, 6, 8, and 10 quanta for other ones.

We examined the results in three series: under constant lighting, under varying illumination conditions, and on a black background under variable illumination. In Table 1 we present the rates of successful localization for each of the test objects. Plots of the overall rates obtained for color spaces using respective distance measures are shown in Fig. 7.[4]

[4] Additional figures can be found on the accompanying CD–ROM.

Obj.	Exp.	found	1	2	Obj.	Exp.	found	1	2
1	4737	40.08%	25.64%	5.65%	9	4594	50.28%	37.61%	4.65%
2	4729	46.56%	30.53%	7.46%	10	4697	34.46%	15.90%	6.77%
3	4720	15.72%	7.41%	3.05%	11	4597	32.67%	13.26%	8.15%
4	4711	50.37%	28.06%	9.17%	12	4701	46.05%	20.71%	10.21%
5	4728	8.79%	2.41%	1.90%	13	4599	19.15%	10.02%	3.30%
6	4719	20.87%	8.58%	3.36%	14	4703	24.21%	10.73%	4.27%
7	4724	21.25%	9.39%	3.25%	15	4568	46.71%	27.49%	8.12%
8	4716	26.78%	12.53%	4.26%	16	4699	8.17%	2.95%	1.57%
					17	4696	60.43%	44.01%	7.21%

Table 1: Results for the individual objects. First column shows the total number of experiments. Second column; the total localization rate. The other columns; the rates for the first and second hypotheses.

7 Conclusions

For Histogram Intersection and the AB color space using 12 quanta, we observed a strong local maximum of the localization rate. This configuration gave the best overall results for variable illumination.

The combination of bin–by–bin distance measures (Histogram Intersection, SSD and χ^2) with the AB color space gave outstanding results. Furthermore, the changes in illumination did not considerably deteriorate the localization rates using AB color space.

Curiously, the proposed RG2 model outperformed the other color spaces in localizing objects using the EMD measure.

The 2–D spaces were as good or better then their 3–D equivalents at localizing objects in real–world scenes (on an arbitrary background). Nevertheless, 3–D spaces performed better on a black background; for Histogram Intersection, recognition rates up to 89% could be realized.

Searching through the image can be dramatically accelerated by applying the lower bound estimations to distance measures. Those methods gave a great performance gain; on average only 13% of all the distances had to be computed.

Naturally, objects were localized better if their sizes in the scene were larger. The evaluation showed, however, that even for small objects satisfactory rates were achieved.

This contribution is an excerpt from the diploma thesis. For details see `http://www.olimp.com.pl/krzyho`.

References

[Ahl98] U. Ahlrichs, B. Heigl, D. Paulus, and H. Niemann. Wissensbasierte aktive Bildanalyse. In *Von der Informatik zu Computational Science und Computational Engineering. Abschlußkolloquium des Sonderforschungsbereichs 182 Multiprozessor– und Netzwerkkonfigurationen*, pages 97–113, Erlangen, 1998. Institut für Mathematische Maschinen und Datenverarbeitung (Informatik).

[Fol92] J. D. Foley, A. van Dam, S. Feiner, and J. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, 1992.

[Pre88] W.H. Press, B.P. Flannery, S. Teukolsky, and W.T. Vetterling. *Numerical Recipes - the Art of Numerical Computing, C Version*. Cambridge University Press, 1988.

[Puz98] J. Puzicha, Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. In H. Burkhard and B. Neumann, editors, *Computer Vision — ECCV '98*, number 1406 in Lecture Notes in Computer Science, pages 563–577. Springer, Heidelberg, 1998.

[Rub98] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In *Proceedings of the 6^{th} International Conference on Computer Vision (ICCV)*, pages 59–66, Bombay, January 1998. IEEE Computer Society Press.

[Sch97] B. Schiele. *Object Recognition using Multidimensional Receptive Field Histograms (English translation)*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble Cedex, 1997.

[Swa91] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, November 1991.

[Vin98] V.V. Vinod and H. Murase. Image retrieval using efficient local area matching. *Machine Vision and Applications*, 11(1):7–15, 1998.

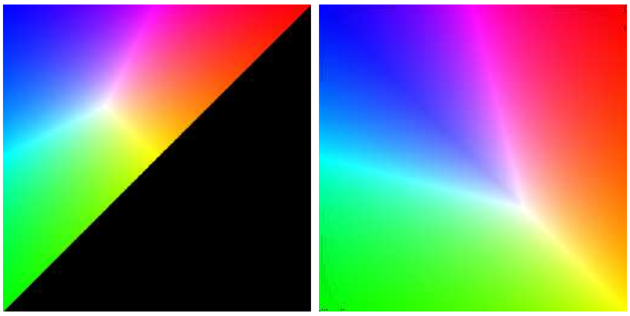

Fig. 2: Normalized RG (left) and RG2 space (right)

Fig. 3: Exemplary test scene

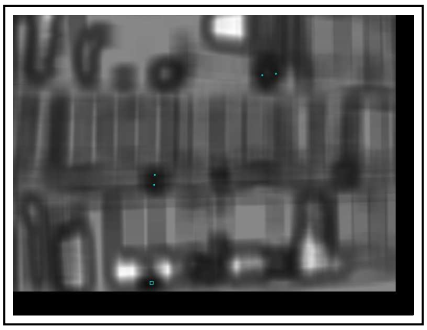

Fig. 4: Interest map for object 11 using SSD measure

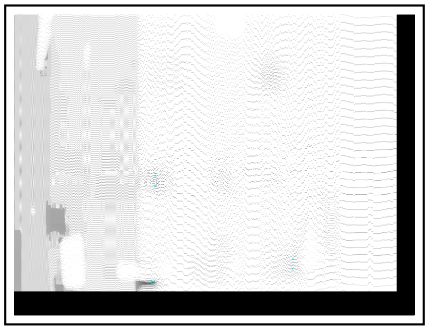

Fig. 5: Interest map for object 11 using Histogram Intersection with Active Search

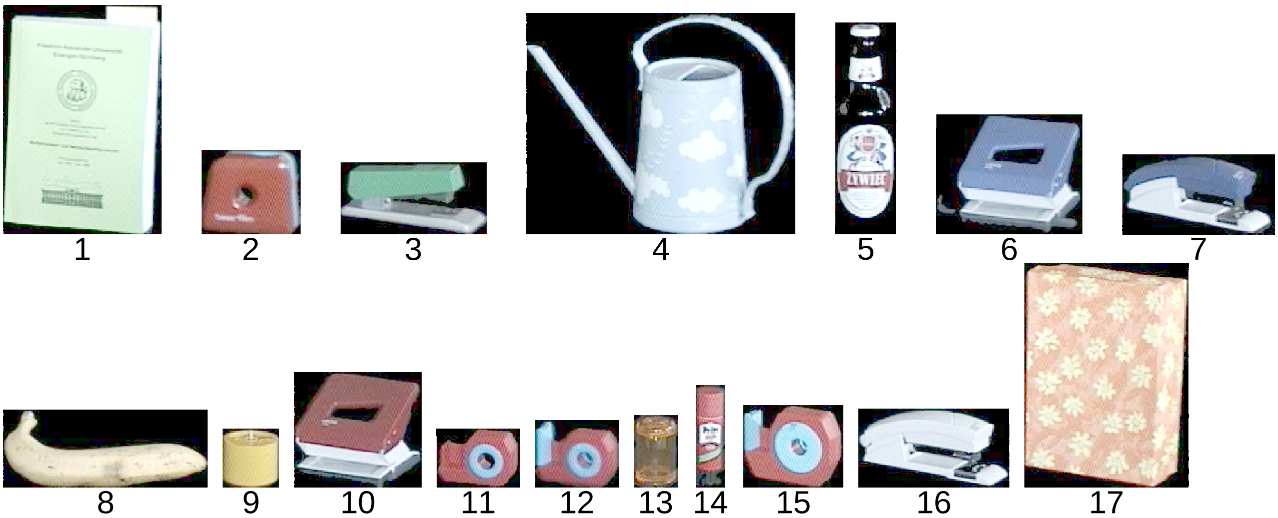

Fig. 6: Data set of the test objects

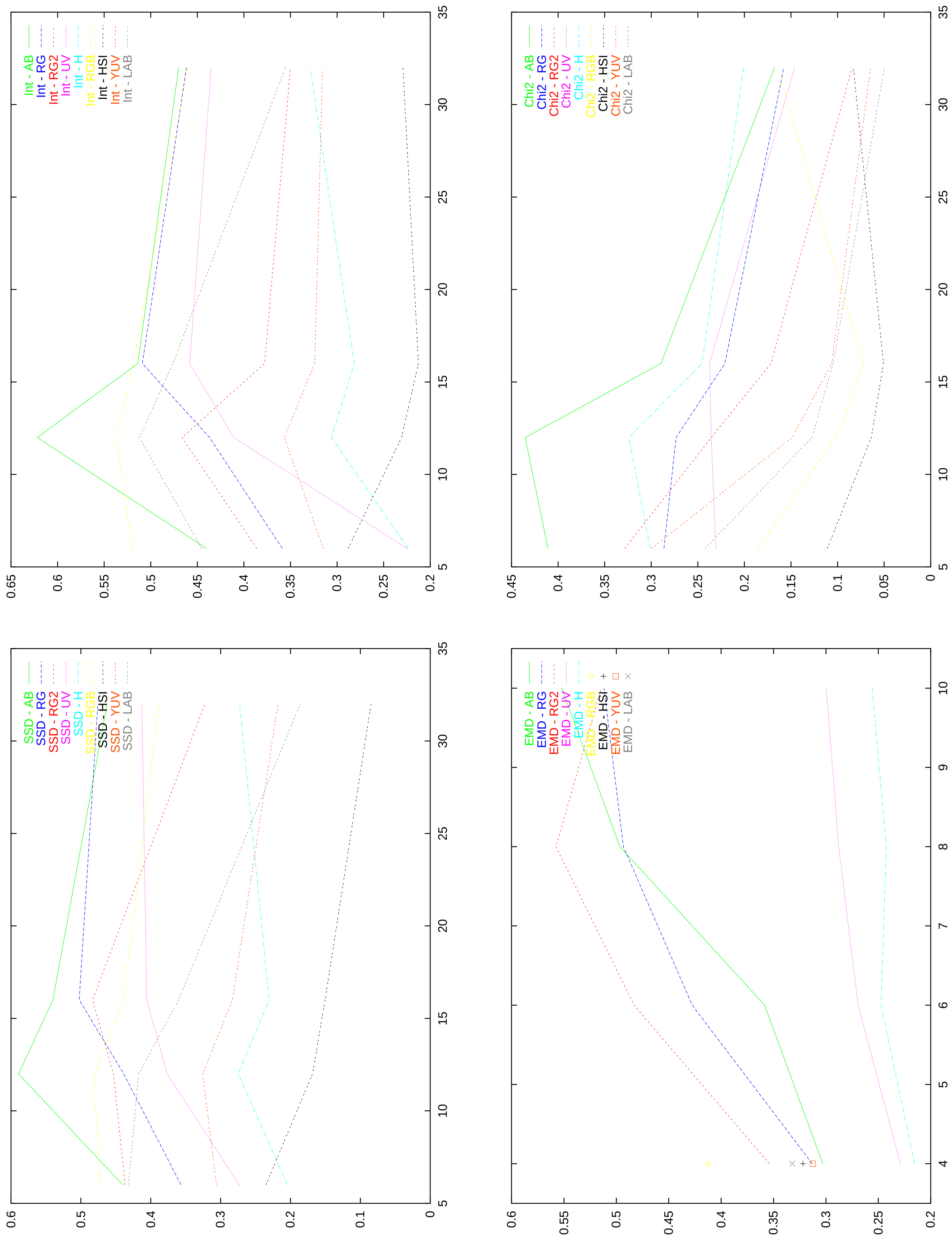

Fig. 7: Localization rates for color space and distance measure combinations under variable illumination conditions: Histogram intersection (top left), χ^2 (top right), SSD (bottom left), EMD (bottom right). The graphs show localization rates versus quantization (number of bins in the histogram).