CSDD Features: Center-Surround Distribution Distance for Feature Extraction and Matching

Robert T. Collins[1] and Weina Ge[1]

The Pennsylvania State University, University Park, PA 16802, USA
{rcollins,ge}@cse.psu.edu

Abstract. We present an interest region operator and feature descriptor called Center-Surround Distribution Distance (CSDD) that is based on comparing feature distributions between a central foreground region and a surrounding ring of background pixels. In addition to finding the usual light(dark) blobs surrounded by a dark(light) background, CSDD also detects blobs with arbitrary color distribution that "stand out" perceptually because they look different from the background. A proof-of-concept implementation using an isotropic scale-space extracts feature descriptors that are invariant to image rotation and covariant with change of scale. Detection repeatability is evaluated and compared with other state-of-the-art approaches using a standard dataset, while use of CSDD features for image registration is demonstrated within a RANSAC procedure for affine image matching.

1 Introduction

One of the key challenges in object recognition and wide-baseline stereo matching is detecting corresponding image regions across large changes in viewpoint. Natural images tend to be piecewise smooth, and most small image regions are therefore near-uniform or edge-like; ill-suited for accurate localization and matching. Although larger regions tend to be more discriminative, they are more likely to span across object boundaries and are harder to match because view variation changes their appearance.

The idea behind interest region detection is to find patches that can be reliably detected and localized across large changes in viewpoint. State-of-the-art approaches include regions that self-adapt in shape to be covariant to image transformations induced by changing rotation, scale and viewing angle [1]. To date, the majority of interest region detectors search for image areas where local intensity structure is corner-like (containing gradients of several different orientations) or blob-like (exhibiting a center-surround contrast difference). Since it is not possible to predict apriori the spatial extent of interesting image structures, detection is often performed within a multiresolution framework where the center and size of interesting regions are found by seeking local extrema of a scale-space interest operator.

Our work is motivated by the goal of finding larger interest regions that are more complex in appearance and thus more discriminative. We seek to improve the utility of larger image regions by using a feature descriptor that is

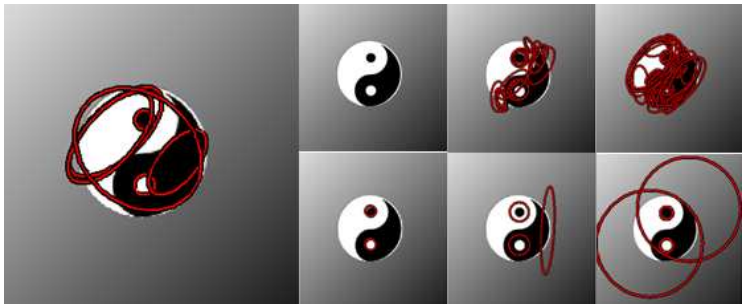

Fig. 1. Detection results for a yin-yang symbol. The leftmost picture shows results from CSDD region detection, which captures the hierarchical composition of the object by correctly finding blobs at the three characteristic scales. The smaller images at right show, from top left to bottom right, the original image and results from five other detectors: Harris-affine, Hessian-affine, IBR, MSER and the Salient region detector. Only the CSDD detector captures the natural location and scale of the symbol.

insensitive to geometric deformation. Specifically, we develop an interest region operator and feature descriptor called Center-Surround Distribution Distance (CSDD) based on comparing empirical cumulative distributions of color or texture extracted from a central foreground region and a surrounding ring of background pixels. The center-surround nature of the approach makes it a variant of blob detection. However, in comparison to typical center-surround schemes that measure difference in average feature contrast, our approach is based on fine-grained histograms of the feature distributions in the central and surrounding regions, compared using a distance measure that takes into account not only the intersection of mass in overlapping bins but also the *ground-distance* between non-overlapping bins.

Our long-term goal is to develop a method capable of extracting entire objects as interest regions, since we believe that this kind of figure-ground separation is needed to make progress in extremely wide-baseline matching scenarios. Although we cannot rely on objects and their surrounding background as having uniform intensity, the distribution of color and texture on an object often differs from that of the background, and thus a center-surround approach to blob detection using feature distributions should have applications to figure-ground segmentation in addition to interest region detection (Fig. 1).

Related Work

A vast literature exists on local feature detection. Of particular interest are detectors that generate features covariant to image transformations [2–7]. Lindeberg studied scale covariant features extensively in his seminal work on scale-space theory [8]. Recent work considers affine covariance to cope with broader types of image deformations. Mikolajczyk and Schmid propose affine covariant detectors based on local image statistics characterized by Harris and Hessian matrices [6]. Tuytelaars and Van Gool incorporate edge information into the

local image structure around corners to exploit geometric constraints that are consistent across viewpoint [4]. Matas et al. extract blob-like maximally stable extremal regions (MSER) whose borders remain relatively unchanged over large ranges of greyscale thresholding [5]. Bay et al. use integral images to build a fast and effective interest point detector and descriptor [7].

Kadir et.al. define a region saliency score composed of two factors: Shannon entropy of local intensity histograms, and magnitude change of intensity histograms across scale. Scale-space maxima of the entropy term identify regions with complex, and thus distinctive, greyscale structure. The inter-scale term serves to both downweight salience at edges, as well as promote the salience of blob-like regions. Indeed, if implemented as a finite difference approximation, this term becomes L1-distance of histograms across scale, which can be interpreted as a center-surround difference operator applied to multi-channel data. Similar center-surround ideas have long been used in salience detection. Itti and Koch [9] apply a Difference-of-Gaussian filter to each channel of a feature map to mimic receptive cell response and salience detection in human vision. More recently [10], color histograms extracted within rectangular central and surrounding regions are compared using the χ^2 distance as one component to measure the salience of the central region.

To find a distinctive representation for detected interest regions, various region descriptors have been developed. Histograms are a simple yet effective representation. Histograms of intensity, color, gradient of intensities [11], and other filter responses are prevalent in vision. Refer to [12] for an extensive discussion on measuring the difference between two histograms.

2 CSDD Features

This section presents an overview of our proposed CSDD feature for interest region extraction. An implementation-level description is provided in Sect. 3.

2.1 Center-Surround Distributions

Given a random variable f defined over a (possibly real and multi-valued) feature space, we define its empirical cumulative distribution function (hereafter referred to simply as a distribution function) within some 2D region of support of the image I as

$$F(v) = \int_{x \in R} \delta(I(x) \leq v)\ w(x)dx \ \ / \ \ \int_{x \in R} w(x)dx \tag{1}$$

where $\delta(\cdot)$ is an indicator function that returns 1 when the boolean expression evaluates to true, and 0 otherwise, R is the real plane, and $w(x)$ is a spatial weighting function defining the spatial extent of the region of support while perhaps emphasizing some spatial locations more than others. For example, a rectangular region with uniform weighting can be defined by

$$w(x, y) = \delta(a \leq x \leq b) \cdot \delta(c \leq y \leq d), \tag{2}$$

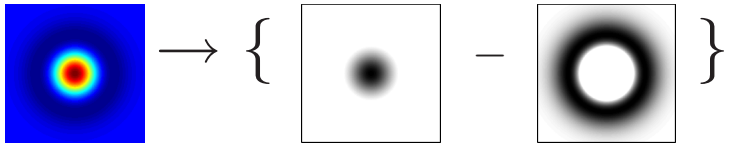

Fig. 2. Weight functions for accumulating center-surround distributions are formed by decomposing the Laplacian of Gaussian $\nabla^2 G(x; x_0, \sigma)$ into weight functions $w_C(x; x_0, \sigma)$ defining a central circular region of radius $\sqrt{2}\sigma$ and $w_S(x; x_0, \sigma)$ defining a surrounding annular region. Specifically, $w_C - w_S = \nabla^2 G$.

whereas

$$w(x) = \exp\{-\frac{(x - x_0)^t (x - x_0)}{2\sigma^2}\}$$

specifies a Gaussian-weighted region centered at point x_0 with circular support of roughly 3σ pixels radius. Now, consider two regions of pixels comprised of a compact central region C and a surrounding ring of neighboring pixels S. Fig. 2 illustrates the center-surround regions we use in this paper. The center region is circular, and the surround region is an annular ring. These regions are defined by weighting functions based on an isotropic Laplacian of Gaussian, $\nabla^2 G(x_0, \sigma)$, with center location x_0 and scale parameter σ. Specifically, letting $r = \|(x - x_0)\|$ be a distance from the center location x_0, the center and surround weighting functions, denoted w_C and w_S, respectively, are defined as:

$$w_C(x; x_0, \sigma) \;=\; \begin{cases} \frac{1}{\pi\sigma^4}\left(1 - \frac{r^2}{2\sigma^2}\right)\; e^{-r^2/2\sigma^2} & r \le \sqrt{2}\,\sigma \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$w_S(x; x_0, \sigma) \;=\; \begin{cases} -\frac{1}{\pi\sigma^4}\left(1 - \frac{r^2}{2\sigma^2}\right)\; e^{-r^2/2\sigma^2} & r > \sqrt{2}\,\sigma \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

These nonnegative weighting functions coincide with the positive and negative *channels* of the LoG function, with the positive channel being the central portion of the "Mexican hat" operator, and the negative channel being the annular ring around it. Thus $w_C - w_S = \nabla^2 G$. Since the LoG operator integrates to zero, we know that the positive and negative channels have equal weight. Integrating in polar coordinates, we find that

$$\iint w_C(\cdot)\; r\, dr\, d\theta \;=\; \iint w_S(\cdot)\; r\, dr\, d\theta \;=\; \frac{2}{e\;\sigma^2} \tag{5}$$

This term $2/(e\,\sigma^2)$ becomes the normalization factor in the denominator of (1) when we extract feature distributions using w_C and w_S.

Why We Don't Use Integral Histograms: Integral histograms are often used to extract feature histograms of rectangular regions efficiently [13]. We choose not to use them for the following reasons:

• Integral histograms require space proportional to storing a full histogram at each pixel. Therefore, in practice, integral histograms are only used for coarsely quantized spaces. We use finely quantized feature spaces where histograms have hundreds of bins, making the integral histogram data structure very expensive.

• Generating rectangular center-surround regions at multiple scales using integral histograms can be viewed as equivalent to generating a scale space by smoothing with a uniform box filter. It is well-known that smoothing with a box filter can lead to high-frequency artifacts. We prefer to use LoG filtering to generate a scale-space that is well-behaved with respect to non-creation/non-enhancement of extrema [8].

• Integral histograms lead to rectangular center-surround regions, aligned with the image pixel array axes. They are not a good choice for rotationally-invariant processing. Our approach yields circular center-surround regions that are rotationally invariant, and a minor modification yields anisotropic elliptical regions (see Sect. 4.1).

2.2 Center-Surround Distribution Distance

Consider two feature distributions F and G computed over the center and surround regions described above. Intuitively, if the feature distributions of F and G are very similar, it may be hard to visually discriminate the central region from its surrounding neighborhood, and this central region is therefore a bad candidate to use as a blob feature. On the other hand, if distributions F and G are very different, the central region is likely to be visually distinct from its surroundings, and thus easy to locate. Our hypothesis is that regions that look different from their surroundings are easy to detect and match in new images of the scene taken from a different viewpoint.

Many dissimilarity measures are available to compare two cumulative distributions F and G, or density functions dF and dG. In practice, information theoretic measures such as χ^2 distance or Bhattacharyya coefficient for comparing two density functions are problematic since they only take into account the intersection of probability mass, not the distance between masses [14]. For example, using these measures to compare greyscale density functions represented as histograms with 256 bins, an image patch with uniform value 0 looks as different from a patch with uniform value 1 as it does from a patch with uniform value 255. This makes such measures unsuitable for use when viewpoint or lighting variation cause shifts in the underlying intensity/color of corresponding image patches. Kolmogorov-Smirnov or Kuiper's test [12] based on comparing cumulative distribution functions F and G are more robust in this regard, as well as being applicable to unbinned data. In this paper, we use Mallow's distance, also known as Wasserstein's distance, and in the present context equiva-

lent to the Earth Mover's Distance (EMD) [15]. Mallow's distance between two d-dimensional distributions F and G can be defined as

$$M_p(F,G) = \min_J \left\{ (E_J \|X-Y\|^p)^{\frac{1}{p}} : (X,Y) \sim J, X \sim F, Y \sim G \right\} \quad (6)$$

where the minimum is taken over all $d \times d$ joint distributions J such that the d dimensional marginal distribution wrt X is F, and the marginal wrt Y is G. In general, finding this minimum is equivalent to solving the Monge-Kantorovich *transportation problem* via, for example, the simplex algorithm [14, 16].

In this paper we use the known but still surprising result that, despite it's intractability in general dimensions, for 1D distributions the transportation problem is solved as

$$M_p(F,G) \;=\; \left(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^p \, dt \right)^{\frac{1}{p}} \quad (7)$$

and that, for $p = 1$, there is a closed form solution that is the L_1 distance between cumulative distributions F and G [15, 17] :

$$M_1(F,G) \;=\; \int_{-\infty}^{\infty} |F(v) - G(v)| \, dv \quad (8)$$

Although restricting ourselves to 1D distributions seems to be a limitation, in practice it comes down to approximating a joint distribution by a set of 1D marginals. We take care to first transform the original joint feature space into an uncorrelated one before taking the marginals (see next section), and also note results from the literature that show the use of marginals outperforming full joint feature spaces when empirical sample sizes are small [15]. We feel that any drawbacks are more than made up for by the ability to use finely quantized marginal distributions while still computing an efficient, closed-form solution.

3 An Implementation of CSDD

We now describe a specific practical implementation of CSDD feature extraction. Like many current detectors, the CSDD detector operates over a scale-space formed by a discrete set of scales. At each scale level, center-surround distributions are extracted at each pixel and compared using Mallow's distance to form a CSDD measure. The larger the CSDD value, the more dissimilar the center region is from its surrounding neighborhood. We therefore can interpret these dissimilarity scores within the 3D volume of space (center pixel) and scale (sigma) as values from a scale-space interest operator (see Fig. 3). Similar to other detectors, we extract an interest region for each point in the volume where the interest function achieves a local maximum across both space and scale, as determined by non-maximum suppression with a 5x5x3 (x, y, σ) neighborhood. Each maxima yields the center point (x, y) and size σ of a center-surround region that we can expect to detect reliably in new images. We choose not to do spatial subsampling at higher scales to form a pyramid, but instead keep the same

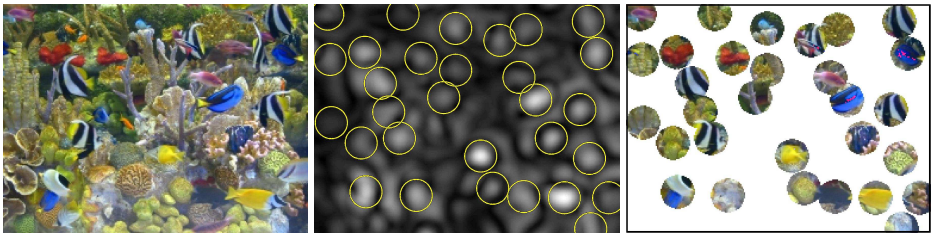

Fig. 3. Sample regions extracted as local maxima of the CSDD interest operator at one scale level. Left: Original color image. Middle: CSDD interest score computed for each pixel, over center-surround regions of scale $\sigma = 9.5$ pixels. Overlaid are the 30 most dominant peaks at this scale level, displayed as circles with radius $\sqrt{2}\sigma$. Right: Interest regions identified by these peaks.

number of pixels at all scale levels. Therefore, features at all scales are spatially localized with respect to the resolution of the original pixel grid, with no need for spatial interpolation.

What remains to be seen is how to generate the scale-space volume of CSDD interest values efficiently. As discussed in the last section, Mallow's/EMD distance can be solved in closed-form if we are willing to approximate a joint RGB distribution using three 1D marginals. To make approximation by marginals better justified, we first transform the color space into a new space where the three random variables representing the color axes are roughly uncorrelated. A simple linear transformation of RGB color space due to Ohta [18] yields a set of color planes that are approximately uncorrelated for natural images. The transformation is $I_1 = (R + G + B)/3$, $I'_2 = R - B$, $I'_3 = (2G - R - B)/2$. Although the Jacobian of this transformation is one, individual axes are scaled non-uniformly so that color information is slightly emphasized (expanded).

We implement the integral in (8) as the sum over a discrete set of sampled color values. Unlike typical histogram-based approaches, we avoid coarse quantization of the feature space or adaptive color clustering. Instead, we sample values finely along each of the marginal color feature axes. In our current implementation, each axis is quantized into 128 values, yielding a concatenated feature vector of size 384 to represent the color distribution of the center region, and another vector of the same size representing the surround region.

The heart of the distribution distance computation thus involves computing $F(v) - G(v)$, the difference between cumulative distributions F and G for a sampled color space value v. Referring to (1) and (5) for the definition of how each distribution is computed and for the value of the normalization constant associated with our two center-surround weight functions w_C and w_S, and the

fact that, by construction, $w_C - w_S = \nabla^2 G$, we see that

$$F(v) - G(v) = \frac{\int_{R^2} \delta(I(x) \leq v)\, w_C(x) dx}{\int_{R^2} w_C(y) dy} - \frac{\int_{R^2} \delta(I(x) \leq v)\, w_S(x) dx}{\int_{R^2} w_S(y) dy} \quad (9)$$

$$= \frac{e\sigma^2}{2} \int_{R^2} \delta(I(x) \leq v)\, [w_C(x) - w_S(x)]\, dx \quad (10)$$

$$= \frac{e\sigma^2}{2} \int_{R^2} \delta(I(x) \leq v)\, \nabla^2 G(x; x_0, \sigma)\, dx \quad (11)$$

Although we have compressed the notation to save space, distributions F and G are also functions of pixel location x_0 and scale level σ, the center and size of the concentric central and surrounding regions of support. Since we want to explicitly compute a difference value for every pixel, the center-surround difference $F(v) - G(v)$ for all pixels at one scale level becomes convolution of the binary indicator function $\delta(I(x) \leq v)$ with a LoG filter of scale σ.

An important implementation detail is how to perform efficient LoG filtering, particularly since we have to perform it many times, once for each of a set of finely sampled values v and at each scale level σ. Standard convolution with a 2D spatial LoG kernel takes time quadratic in the radius of the support region. While a separable spatial implementation would reduce this to linear time, this is still expensive for large scale levels. Our implementation is based on a set of recursive IIR filters first proposed by Deriche and later improved by Farneback and Westin [19, 20]. These fourth-order IIR filters compute the Gaussian and its first and second derivatives (and through combining these, the LoG operator) with a constant number of floating point operations per pixel, regardless of the spatial size σ of the operator.

Although there is a superficial similarity between our implementation and greyscale blob detection by LoG filtering, note that we are filtering multiple binary masks formed from a fine-grained feature distribution and combining the results into a distance measure, not computing a single greyscale convolution. Our approach measures Mallow's distance between two distributions, not difference between average grey values. It is fair to characterize the difference between our method and traditional LoG blob detection as analogous to using a whole distribution rather than just the mean value to describe a random variable.

4 Experimental Validation

As discussed, we compute a distance between empirical cumulative feature distributions to measure how similar a circular central region is in appearance to its surrounding annular ring of pixels. Our hypothesis is that regions that look markedly different from their surroundings are likely to be good features to use for detection and matching.

One important aspect of an interest region detector's performance is repeatability of the detections. This measures how often the same set of features are detected under different transformations such as changes in viewing angle or

lighting. In Sect. 4.1, we adopt the framework of [1] to compare repeatability of our detector against other state-of-the-art approaches. Moreover, it is also desirable for extracted regions to have high descriptive power. In Sect. 4.2 we demonstrate the utility of CSDD features for correspondence matching under affine image transformation.

4.1 Repeatability experiments

We test the repeatability of our detector using the standard dataset for evaluation of affine covariant region detectors.[1] CSDD regions are detected by the procedure discussed in Sect. 3. To improve the accuracy of scale estimation, for each detection corresponding to a local maximum in the 3D volume of space and scale, we do a 3-point parabolic interpolation of CSDD response across scale to refine the peak scale and response value. We also compute the Hessian matrix of responses around the detected points at each scale level. This matrix of second order derivatives is used for two purposes: 1) to remove responses due to ridge-like features [11]; and 2) to optionally adapt our circular regions to ellipses, using the eigenvectors and eigenvalues of the Hessian matrix to define the orientation and shape of an ellipse constrained to have the same area as the original circular region of support. Although this is a much simpler and coarser ellipse fitting method than the iterative procedure in [1], it does improve the performance of the CSDD detector for large viewpoint changes. We refer to this version of the implementation as the elliptical CSDD (eCSDD), to distinguish it from the original circular CSDD (cCSDD). We also set a conservative threshold on CSDD response score to filter out very weak responses.

The evaluation test set consists of eight real world image sets, generated by image transformations with increasing levels of distortion, such as view angle, zoom, image blur, and JPEG compression. For each reference and transformed image pair, a repeatability score is computed as described in [1]. Fig. 5 compares repeatability results from our CSDD detector against the five detectors evaluated in [1]. We outperform the state of the art in three out of the eight test cases (the orange line in Fig. 5b,c,d) and achieve comparable results for the others, even though our detector is designed to handle only isotropic scale and rotation changes. Two of the three cases where we outperform all other detectors are the ones that test ability to handle large amounts of zoom and rotation. For cases where viewpoint changes induce a large amount of perspective foreshortening, elliptical adaptation improves the performance (Fig. 5a).

We show elliptical CSDD regions side-by-side with regions from the best-performing state-of-the-art comparison detector for two sample cases in Fig. 4. In both cases, the CSDD regions are able to capture the meaningful structures in the scene such as the green leaves in the bark image and the windows on the boat. Our region density is of relatively similar order as MSER, EBR, and IBR detectors. For more details on the region density and the computational cost of CSDD feature detection, please refer to our website [2]. We did not include the

[1] http://www.robots.ox.ac.uk/∼vgg/research/affine/ as of March 2008.
[2] http://vision.cse.psu.edu/projects/csdd/csdd.html

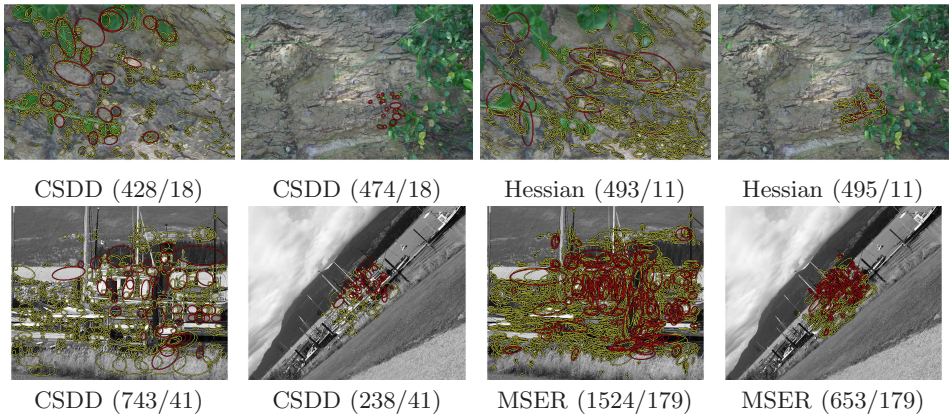

Fig. 4. Regions found by CSDD (*left*) and the comparison detector (*right*) for two pairs of test images. The detected regions that fall within the common image area are in yellow. The corresponding regions are in red. Each image is labeled with the total number of detections within the image versus the number of correspondences found in both of the images. **Top**: Regions found in bark images differing by a scale factor of 4. **Bottom**: Regions found in boat images differing by a scale factor of 2.8.

performance curve of the salient region detector [2] because of its absence in the online evaluation package and its consistently low rank for most of the cases in this repeatability test, according to [1].

4.2 Matching experiments

Ultimately, the best test of a feature detector/descriptor is whether it can be used in practice for correspondence matching. We ran a second series of experiments to test the use of CSDD features to find correspondence matches for image registration. To demonstrate the robustness and the discriminativeness of the original circular CSDD regions, we do not include the elliptical adjustment procedure in this matching test. A simple baseline algorithm for planar image registration was used, similar to the one in [21]. The method estimates a 6-parameter affine transformation matrix based on RANSAC matching of a sparse set of feature descriptors. In [21], corner features are detected in each frame, and a simplified version of the linear assignment problem (aka marriage problem) is solved by finding pairs of features across the two images that mutually prefer each other as their best match, as measured by normalized cross-correlation of 11x11 intensity patches centered at their respective locations. This initial set of candidate matches is then provided to a RANSAC procedure to find the largest inlier set consistent with an affine transformation. We modify this basic algorithm by replacing corner features with our CSDD features, and replacing image patch NCC with average Mallow's distance between the two center distributions and

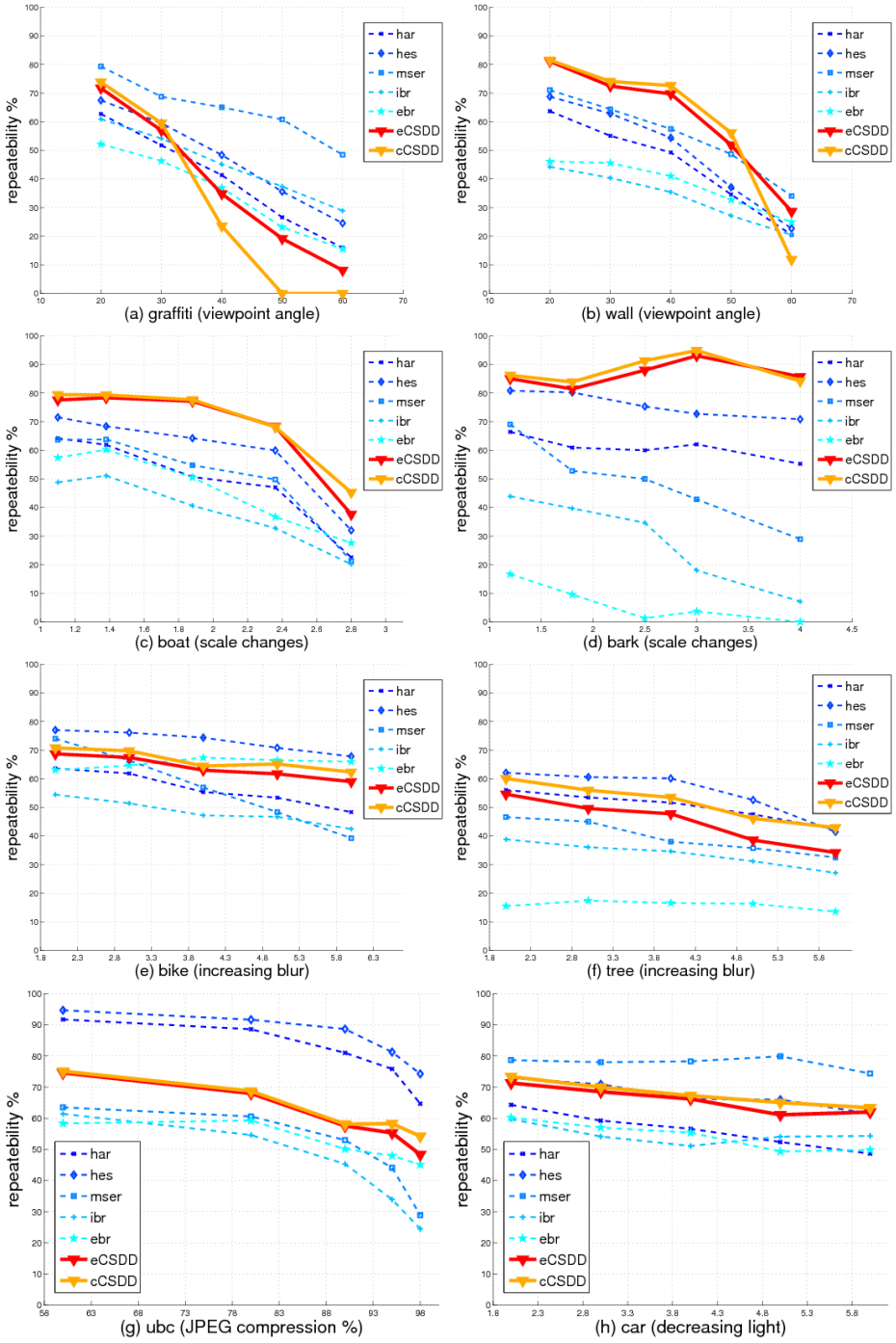

Fig. 5. Comparison of repeatability scores between the circular cCSDD detector, the elliptical eCSDD detector, and five other state-of-the-art detectors (the Harris- and Hessian-affine detectors, the MSER detector, the edge-based detector (EBR), and the intensity extrema-based detector (IBR)) for the eight image sequences from the standard evaluation dataset[1].

Fig. 6. Affine Ransac Image Matching Experiments. **Row 1**: four frames from a parking lot video sequence, showing affine alignment of bottom frame overlaid on top frame. The full video sequence of results is provided on our website. **Row 2**: left to right: shout3 to shout4; shout2 to was2 (images courtesy of Tinne Tuytelaars); stop sign; snowy stop sign. **Row 3**: kampa1 to kampa4 (images courtesy of Jiri Matas); bike1 to bike6; trees1 to trees5; ubc1 to ubc6. **Row 4**: natural textures: asphalt; grass; gravel; stones.

two surround distributions of a pair of interest regions. Note that use of CSDD features is a vast improvement over single-scale (11x11) corner patches in terms of being able to handle arbitrary rotation and scaling transformations.

Sample results of this baseline matching algorithm are shown in Fig. 6. Higher resolution versions of these pictures along with more examples and a complete video sequence of matching results on the parking lot sequence are provided on our website [2]. The use of circular CSDD regions limits our performance under out-of-plane rotations. Although we can get more matching features on tilted planes if we use elliptical adaptation and relax the inlier distance threshold, images with large amounts of planar perspective foreshortening should be handled with a homography-based match procedure.

5 Discussion

We have presented a new scale-space interest operator based on center-surround distribution distance (CSDD). The method finds rotationally invariant and scale covariant interest regions by looking for locations and scales where empirical cumulative distributions between a central region and its surrounding local neighborhood are dissimilar. The proposed approach performs competitively on the standard evaluation test for affine covariant region detectors, where it outperforms the state-of-the-art in three out of the eight scenarios. We have also tested the use of CSDD as a feature descriptor within a RANSAC-based affine image matching framework. In our experience, we find that CSDD feature extraction and matching works well on textured, natural images, and performs very well under large changes of scale and in-plane rotation.

While the use of Mallow's/EMD distance is resilient to some changes in lighting, large illumination changes currently defeat our baseline matching algorithm. This is partly due to the simplified marriage problem solution of only picking pairs of candidate features that are mutually each other's best match. With large changes in lighting, it is often the case that some other feature patch becomes a better match in terms of intensity/color than the correct correspondence. Matching based only on color distributions, without taking into account any spatial pattern information, is a brittle approach, and our results can be viewed as surprisingly good in this regard. Our ability to match at all, using only "color histograms", is due to the fact that we finely quantize the color spaces, and use a robust measure of color similarity. However, a practical system would also need to incorporate additional feature descriptor information, such as SIFT keys computed from normalized patches defined by CSDD spatial support regions [22].

References

1. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. International Journal of Computer Vision **65** (2005) 43–72

2. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: European Conference on Computer Vision. (2004) 345–457
3. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets, or "How do i organize my holiday snaps?". In: European Conference on Computer Vision. (2002) 414–431
4. Tuytelaars, T., Gool, L.V.: Content-based image retrieval based on local affinely invariant regions. In: International Conference on Visual Information Systems. (1999) 493–500
5. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference. (2002) 384–393
6. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: European Conference on Computer Vision. (2002) 128–142
7. H. Bay, T.T., Gool, L.V.: Surf: Speeded up robust features. In: European Conference on Computer Vision. (2006) 404–417
8. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer, The Netherlands (1994)
9. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews Neuroscience **2** (2001) 194–203
10. Liu, T., Sun, J., Zheng, N.N., Tang, X., Shum, H.Y.: Learning to detect a salient object. In: IEEE Computer Vision and Pattern Recognition. (2007) 1–8
11. Lowe, D.: Distinctive image features from scale-invariant keypoints, cascade filtering approach. International Journal of Computer Vision **60** (2004) 91–110
12. Rubner, Y., Puzicha, J., Tomasi, C., Buhmann, J.: Empirical evaluation of dissimilarity measures for color and texture. Computer Vision and Image Understanding **84** (2001) 25–43
13. Porikli, F.: Integral histogram: A fast way to extract higtograms in cartesian spaces. In: IEEE Computer Vision and Pattern Recognition. (2005) I: 829–836
14. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40** (2000) 91–121
15. Levina, E., Bickel, P.: The earth mover's distance is the Mallows distance: Some insights from statistics. In: International Conference on Computer Vision. (2001) II: 251–256
16. Q. Zhao, S.B., Tao, H.: Differential emd tracking. In: International Conference on Computer Vision. (2007) 1–8
17. Villani, C.: Topics in Optimal Transportation. Volume 58. Graduate Studies in Mathematics series, American Mathematical Society (2003)
18. Ohta, Y., Kanade, T., Sakai, T.: Color information for region segmentation. Computer Graphics and Image Processing **13** (1980) 222–241
19. Deriche, R.: Fast algorithms for low-level vision. IEEE Trans. Pattern Analysis and Machine Intelligence **12** (1990) 78–87
20. Farneback, G., Westin, C.F.: Improving Deriche-style recursive gaussian filters. Journal of Mathematical Imaging and Vision **26** (2006) 293–299
21. D.Nister, Naroditsky, O., Bergen, J.: Visual odometry. In: IEEE Computer Vision and Pattern Recognition. (2004) I: 652–659
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Analysis and Machine Intelligence **27** (2005) 1615–1630