Best-Buddies Similarity for Robust Template Matching

Tali Dekel[1], Shaul Oron[2], Michael Rubinstein[3], Shai Avidan[2], William T. Freeman[1]
[1]MIT CSAIL [2]Tel Aviv University [3]Google Research

We propose a novel method for template matching in unconstrained environments. Its essence is the Best Buddies Similarity (BBS), a useful, robust, and parameter-free similarity measure between two sets of points.

Template matching methods have been used with great success over the years but they still suffer from a number of drawbacks. Typically, all pixels (or features) within the template and a candidate window in the target image are taken into account when measuring their similarity. This is undesirable in some cases, for example, when the background behind the object of interest changes between the template and the target image (see Fig.). In such cases, the dissimilarities between pixels from different backgrounds may be arbitrary, and accounting for them may lead to false detections of the template (see Fig. (b)).

In addition, many template matching methods assume a specific parametric deformation model between the template and the target image (e.g., rigid, affine transformation, etc.). This limits the type of scenes that can be handled, and may require estimating a large number of parameters when complex deformations are considered.

Our method addresses these problems, and thus can be applied successfully to template matching *in the wild*. In the paper, we introduce the BBS, analyze its key features, and perform extensive evaluation of its performance compared to a number of commonly used alternatives on a challenging dataset.

Best-Buddies Similarity: BBS measures the similarity between two sets of points $P=\{p_i\}_{i=1}^N$ and $Q=\{q_i\}_{i=1}^M$, where $p_i, q_i \in \mathbb{R}^d$. A key feature of this measure is that it relies only on a subset (usually small) of pairs of points – the *Best Buddies Pairs (BBPs)*. A pair of points is considered a BBP if each point is the nearest neighbor of the other in the corresponding point set. BBS is then taken to be the fraction of BBP out of all the points in the set. Formally,

$$bb(p_i,q_j,P,Q) = \begin{cases} 1 & \mathrm{NN}(p_i,Q)=q_j \wedge \mathrm{NN}(q_j,P)=p_i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where, $\mathrm{NN}(p_i,Q)=\underset{q}{\operatorname{argmin}}\, d(p_i,q)$, and $d(p_i,q)$ is some distance measure. The BBS between the point sets P and Q is given by:

$$\mathrm{BBS}(P,Q) = \frac{1}{\min\{\mathrm{M,N}\}} \cdot \sum_{i=1}^{N}\sum_{j=1}^{M} bb(p_i,q_j,P,Q). \tag{2}$$

Albeit simple, BBS has important and nontrivial properties. BBS relies only on a (usually small) subset of matches i.e., pairs of points that are BBPs, whereas the rest are considered as outliers. Furthermore, the BBPs are found without any prior knowledge on the data or its underlying deformation. Another, less obvious property is that the BBS between two point sets is maximal when the points are drawn from the same distribution, and drops sharply as the distance between the distributions increases. In the paper we provide a statistical formulation of this observation, and analyze it numerically in the 1D case. The ability of BBS to reliably match features coming from the same distribution, in the presence of outliers, makes it highly attractive for robust template matching under visual changes and geometric deformations.

To apply BBS to template matching, one needs to convert each image patch to a point set in $\mathbb{R}^d$. To this end, we represent an image window in a spatial-appearance space. That is, we break the region into $k \times k$ distinct patches. Each $k \times k$ patch is represented by a k^2 vector of its *RGB* values and *xy* location of the central pixel, relative to the patch coordinate system. In the paper, we show how the BBS can be efficiently computed between the template and every possible window (of the size of the template) in the image.

In the paper, we perform qualitative as well as extensive quantitative evaluation of our method using over 100 real world examples, and compared it to seven other commonly used template matching methods. Several examples are presented in Fig. and Fig. 2, showing that our method successfully matches the template to the image despite the drastic change in the template appearance (caused by geometric deformation, partial occlusions, and change of background). As can be seen in Fig (c), BBS captures the bidirectional inliers, which are mostly found on the object of interest, and the confidence maps of BBS, presented in Fig. 2(c), show distinct and well-localized modes. The full evaluation, data, and code are publicly available at: *http://people.csail.mit.edu/talidekel/Best-Buddies Similarity.html*

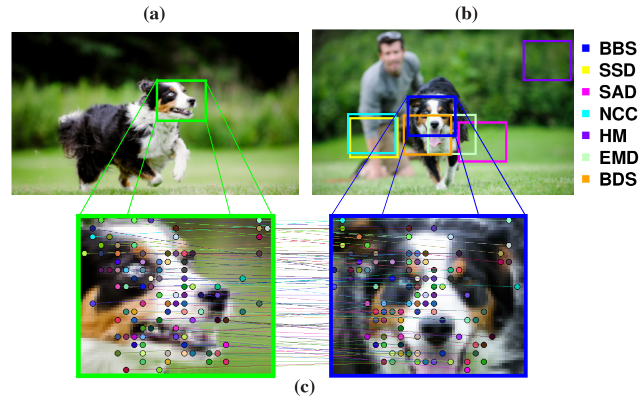

Figure 1: **Best-Buddies Similarity (BBS) for Template Matching:** (a), The template, marked in green, contains an object of interest against a background. (b), The object in the target image undergoes complex deformation (background clutter and large geometric deformation); the detection results using different similarity measures are marked on the image (see legend); our result is marked in blue. (c), The Best-Buddies Pairs (BBPs) between the template and the detected region are mostly found the object of interest and not on the background; each BBP is connected by a line and marked in a unique color.

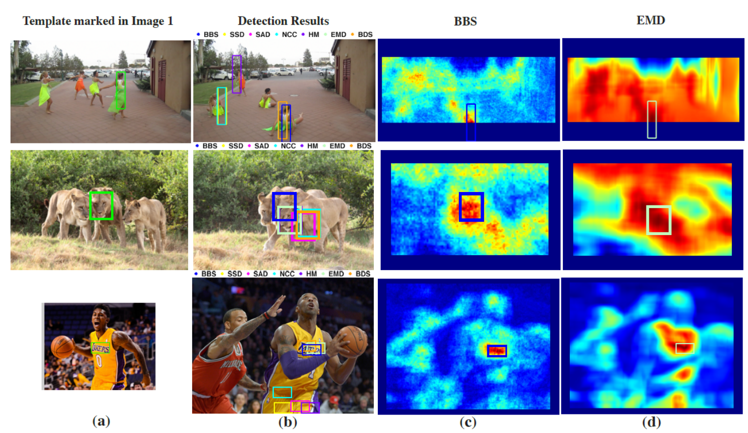

Figure 2: **BBS Template Matching Results on Real Data:** (a), the templates are marked in green over the input images. (b) the target images marked with the detection results of 6 different methods (see text for more details). BBS results are marked in blue. (c)-(d), the resulting likelihood maps using BBS, and EMD , respectively; each map is marked with the detection result, i.e., its global maxima.