

DFlow and DField: New Features for Capturing Object and Image Relationships

Pavel Kisilev, Daniel Freedman, Eugene Wallach, Asaf Tsadok, Yehuda Naveh
IBM Research, Haifa, Israel
{pavelki, danielfr, walach, asaf, naveh}@il.ibm.com

Abstract

In this paper we propose two new types of features useful for problems in which one wants to describe object or image relationships rather than objects or images themselves. The features are based on the notion of distribution flow, as derived from the classic Transportation Problem. Two variants of such features, the Distribution Flow (DFlow) and Displacement Field (DField), are defined and studied. The proposed features show promising results in two different applications, Inter- and Intra-Class Relationship Characterization, and improve on simple concatenation of corresponding pairs of histograms.

1 Introduction

Many computer vision applications deal with object or image descriptors, and their use in classification or recognition tasks. In this paper, we examine an interesting variation on this theme: rather than finding descriptors for individual object or image classes, we define descriptors for *relationships between pairs* of object or image classes. This approach can be quite valuable – understanding the relationship between the paintings of Monet and Toulouse-Lautrec, for example, may tell us something about their respective styles, which would have been difficult to glean directly from separate analyses of their paintings.

The problem of trying to characterize such relationships seems to have been relatively unstudied thus far. The well known work of Hertzmann *et al.* [2] famously examined the notion of “image analogies”; the goal, however, was not to characterize such analogies, but to perform *de novo* image synthesis. Our focus, instead, will be on the characterization of such relationships.

2 The DFlow and DField Features

We begin this section by describing the DFlow, or *Distribution Flow* feature in general terms. On the ba-

sis of this definition, we are then able to define a second feature, called DField, or *Displacement Field*, which is more compact than DFlow but captures its salient elements. We then go on to show how such features may be used in understanding two different types of relationships within computer vision: relationships *between* objects (or images), and relationships *within* objects (or images). Both sets of relationships arise naturally within the study of computer vision, and in Section 3, we give examples of both, and experimental evidence which demonstrates that the new features are useful in characterizing these relationships.

2.1 Definition of DFlow

In general terms, the DFlow feature is designed to capture the relationship between two probability distributions. This is a convenient way to frame the problem of object or image relationships, as both objects and images may be captured by distributions; indeed, this standard technique is quite prevalent in object recognition, for example in the Bag of Visual Words technique (see e.g. [4]). Of course, there are several ways to capture the relationship between distributions. Our solution to this problem is to use the classic Transportation Problem [3, 5] to compute the transformation between the two distributions.

Let us begin with some notation. Pixels in objects or images are characterized by a feature vector $z \in \mathbb{R}^d$; in our case, we choose a combination of 3 color channels and texture (computed for example, as local neighborhood entropy). An object or image, in its entirety, is then characterized by a probability distribution over z . For simplicity, we assume discrete distributions or histograms for the two objects or images, which we write compactly as a list of histogram bins with non-zero probability, i.e.

$$\{(z_i, p_i^k)\}_{i=1}^n$$

where z_i is a bin-center, p_i^k is the corresponding probability mass for that bin for the k^{th} probability distribution ($k = 1, 2$), and n is the number of such bins. Note

that we have fixed both the number of bins and the bin-centers across the two distributions.

The Transportation Problem is then formulated as follows [3, 5]. Let the *DFlow* between the first and second distributions be given by f_{ij} , where the indices i and j range over the (non-empty) bins of the first and second distributions, respectively. That is, f_{ij} can be thought of as the part of bin i from the first distribution which is mapped to bin j of the second distribution. Now, let the *feature distance* between two feature vectors be given by $D(z_1, z_2)$; this may be the ordinary L_2 distance, or it may be some more complex metric. Taking this feature distance as given for the moment, we would like to solve the following optimization:

$$\begin{aligned} \min_{\{f_{ij}\}} & \sum_{i=1}^n \sum_{j=1}^n f_{ij} D(z_i^1, z_j^2) \\ \text{subject to} & \sum_{j=1}^n f_{ij} = p_i^1 \quad i = 1, \dots, n \\ & \sum_{i=1}^n f_{ij} = p_j^2 \quad j = 1, \dots, n \end{aligned}$$

The goal of the objective function is to map the feature vectors from the first distribution z_i^1 to corresponding feature vectors from the second distribution z_j^2 in such a way that the feature distance between them is as small as possible. However, we cannot reasonably expect that each bin of the first distribution maps neatly to exactly one bin of the second distribution. Thus, we allow bins from the first distribution to be spread over several bins from the second distribution, subject to the two constraints which ensure “conservation of probability” for both of the distributions. The Dflow feature f_{ij} thus captures the transformation of one distribution into another.

Before making further comment, we note that the Transportation Problem has been used in computer vision applications before, most notably for the computation of the Earth Mover’s Distance (EMD) [6]. Indeed, a by-product of the EMD computation is the flow that will be useful to us; however, in the context of EMD the flow is nothing more than a by-product, and is used only to facilitate the computation of the EMD metric. By contrast, in the development of our algorithm, the flow itself plays a critical role. It is also worth noting that a few works have used somewhat related ideas [1], albeit in the continuous setting and for different applications, such as registration.

2.2 Definition of DField

The DFlow feature f_{ij} is designed to capture the relationship between distributions. However, one may notice that is a rather large descriptor: its space complexity

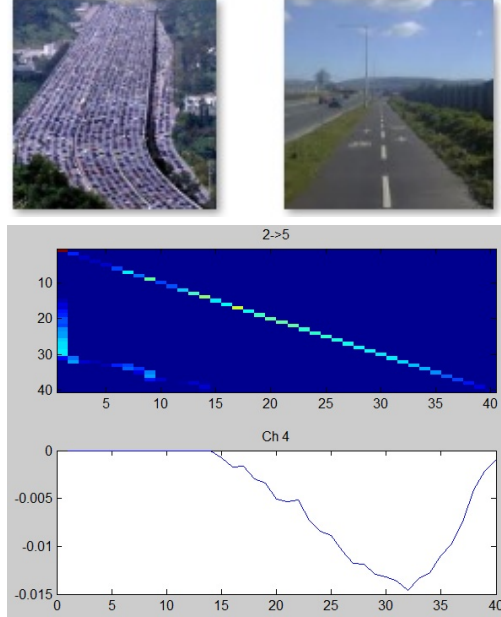


Figure 1. Top: Images of roads with and without traffic. Middle: DFlow descriptor. Bottom: DField descriptor. Both descriptors are computed from texture-based features of the two images.

is $O(n^2)$, where n is the size (number of bins) in each distribution. Furthermore, it is often the case in practice that the DFlow is sparse, with many zero elements. It is natural, therefore, to consider the possibility of a more compact descriptor, based on DFlow.

The DField descriptor captures – for each bin – where its probability mass moves. For bin i of the first distribution, we define the DField δ_i by

$$\delta_i = \sum_j f_{ij} (z_j - z_i)$$

To understand this formula, let us imagine for a moment that for i fixed, f_{ij} is a probability distribution. (In fact, it is not, as $\sum_j f_{ij}$ will generally be smaller than 1.) In this case, $z_j - z_i$ is the displacement bin i undergoes (in feature space) in moving to bin j , so the DField δ_i is like an expected displacement (again, in feature space). This a very useful summary feature – it indicates how the bins of the first distribution must move, in order to transform into the second distribution.

Figure 1 illustrates the DFlow and DField descriptors for two images of roads, with and without traffic. The descriptors were computed from histograms of texture-based features of two images. The DField descriptor reveals a negative dip, which indicates there is “less”

texture in the non-traffic image, as compared to traffic image (as one would expect).

2.3 Inter-Class Relationship Characterization

We now come the first scenario in which we wish to apply our new relationship-based features. In the case of inter-class relationship characterization, the goal is to use the DFlow or DField features in order to describe the relationship between two classes of images. For illustrative purposes, the reader may imagine one class to be the class of paintings by Monet, with the second class the set of paintings by Toulouse-Lautrec. The claim which is put forward in this paper is that the relationship between two paintings, one of Monet and one of Toulouse-Lautrec, may be captured by our new features. In particular, let A be Monet class, and B the Toulouse-Lautrec class. The claim is then as follows:

Claim: Suppose we compute the DFlow $f_{ij}^{AA'}$ between two images of the same class A , and f_{ij}^{AB} between two images of opposite classes A and B . Then $f_{ij}^{AA'}$ and f_{ij}^{AB} will differ significantly. Furthermore, given two training sets of DFlows, $f_{ij}^{AA'}$ and f_{ij}^{AB} , a learning algorithm can learn to classify any new pair as either AA' or AB .

Three comments are in order. First, this claim is an empirical claim, which is borne out in Section 3. Second, one should also include the relationship BB' in any actual experiment; we have done so in Section 3, but have omitted this detail in the above claim, to make the claim clearer. Third, this claim also applies to the DField feature.

2.4 Intra-Class Relationship Characterization

In this second scenario, we wish to use our DFlow and DField descriptors to characterize a single class of objects. How to do this appears less clear than the prior scenario, as the descriptors are meant to capture relationships *between* objects or images. We exploit this relationship-based approach to characterize *feature points* within an object. Based on the presence of these feature points and accompanying DFlow or DField descriptors, one may then use a standard Bag of Visual Words [4] approach to perform object recognition.

The feature points and associated descriptors are computed as follows. An edge detector is run within the image/region of interest. For each sufficiently strong edge-response, we place a circle with center at the edge point, and fixed small radius (e.g. 5-10 pixels). This circle is partitioned into two even halves using the estimated direction of the edge (though other partitioning methods are possible). For each of the two halves of

the circle, $k = 1, 2$, a distribution p_i^k may be computed. The DFlow f_{ij} between the two distributions p_i^1 and p_i^2 is then the descriptor for the feature point. Alternatively, the DField δ_i based on the DFlow f_{ij} may be used as the descriptor.

Given a training set on which one has collected the DFlow or DField descriptors for each feature point, one may then perform vector quantization on the collection of descriptors. If the vector quantization is into L possibilities, then each object in the training set is characterized by an (unnormalized) histogram of size L , based on how many of each type of descriptor exists in the object. This is the Bag of Visual Words approach. Any standard machine learning method for learning classification (e.g. SVM) may then be applied.

3 Applications

In this section we describe two experiments. In the first experiment, we show the ability of the DFlow and DField descriptors to characterize inter-class relationships, as described in Section 2.3. We use three different sets of image class pairs, shown in Figures 1 and 2: (a) paintings of Toulouse-Lautrec vs. paintings of Monet; (b) urban vs. rural scenes; (c) roads with and without traffic. For a given image class pair, the task is as follows: given a pair of images, we wish to classify them as either coming from the same class (e.g. urban-urban or rural-rural), or from opposite classes (e.g. urban-rural). In this sense, we achieve inter-class relationship classification.

To determine the effectiveness of the DFlow and DField descriptors, we compare them with a descriptor which is based on computing the histograms of the two images, and simply concatenating these histograms. We refer to this descriptor as ‘‘Concatenated Histogram’’. In all cases, the image features are Lab combined with two texture-based features; rather than using 5-dimensional histograms, we use five separate 1-dimensional histograms for simplicity (and concatenate the descriptors). For our classification algorithm, we use a simple nearest-neighbor scheme (though SVM or other schemes could easily be substituted), and measure performance, in terms of the correct detection of the given relationship class, using leave-one-out validation.

The results of this experiment are presented in Figure 3. In each case, DFlow achieves the best performance, followed by DField and then Concatenated Histograms. In all cases, the gains in performance yielded by DFlow are impressive: the gain over the Concatenated Histogram feature is more than 20% in each case, and in two cases close to 30%. While DField does not perform quite as well as DFlow, it is still considerably better than the Concatenated Histogram.

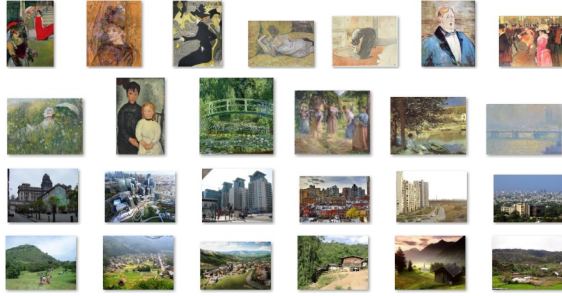


Figure 2. Examples of pairs of image classes. Top rows: paintings of Toulouse-Lautrec (first row) vs. Monet (second row). Bottom rows: urban (third row) vs. rural scenes (fourth row).

In the second experiment, we tested the intra-class object classification of the proposed descriptors, as described in Section 2.4. Here, we applied our method to the problem of identifying skin-tumor-prone freckles (the first class) versus normal freckles (the second class); see Figure 4. Each image contained one or several freckles, and for each descriptor we used a dictionary of 30 Visual Words (with each word being either a DFlow, DField, or a Concatenated Histogram). As in the first experiment, we used a leave-one-out nearest neighbor classifier for simplicity; the classification accuracies are 80.5%, 80%, and 78% for DFlow, DField and Concatenated Histograms, respectively. Once again, DFlow performs best, followed by DField and then Concatenated Histograms; though the gains are more modest than in the first experiment.

4 Conclusions and Future Directions

We have presented DFlow and DField, two variants of a new type of feature based on distribution flow from the Transportation Problem. The proposed features were studied in two different applications, namely, Inter- and Intra-Class Relationship Characterization, and showed promising results. Future directions include exploiting the new features in other computer vision applications, as well as in concert with SVM for potentially better classification performance.

References

- [1] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.
- [2] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin. Image analogies. In *Proceedings of the 28th*

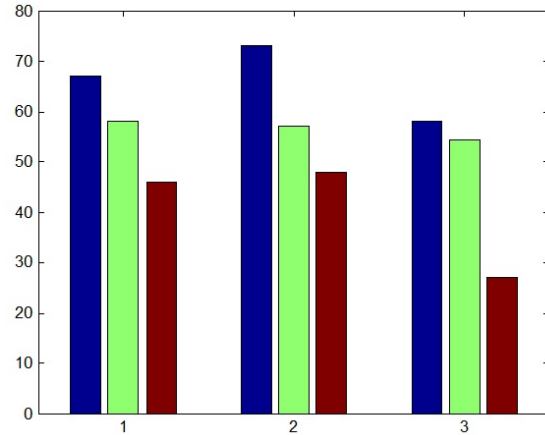


Figure 3. Image inter-class relationship classification accuracy in %. Groups: (1) Monet/Toulouse-Lautrec; (2) urban/rural; (3) traffic/non-traffic. The left (blue) bar is DFlow, the middle (green) is DField, the right (red) is Concatenated Histograms.



Figure 4. Images used in the intra-class relationship experiment. Top: skin-tumor-prone freckles; bottom: normal freckles.

annual conference on Computer graphics and interactive techniques, pages 327–340. ACM, 2001.

- [3] F. Hitchcock. The distribution of a product from several sources to numerous localities. *J. Math. Phys. Mass. Inst. Tech.*, 20:224–230, 1941.
- [4] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. Ieee, 2007.
- [5] S. Rachev. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and its Applications*, 29:647, 1985.
- [6] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.