Enhancing SURF Feature Matching Using Colour Histograms

Tony Marrero Barroso
Centre for Image Processing and Analysis
Dublin City University
tony.marrero@eeng.dcu.ie

Paul F. Whelan
Centre for Image Processing and Analysis
Dublin City University
paul.whelan@dcu.ie

***Abstract*—A strategy is proposed that enhances the local feature matching capabilities of the SURF descriptor by utilising colour histograms. The results compare variations of the RGB, HSV and Opponent colour spaces on a dataset of image pairs that undergo illumination, viewpoint and translational changes. This study finds the most appropriate colour space that enhances the distinctiveness of a descriptor when applied to the matching of corresponding features in arbitrary image sets.**

I. Introduction

Local image features are essential to many computer vision applications such as scene and object Recognition, 3D Reconstruction and Augmented Reality. The task of finding corresponding points can be divided into three main steps: *detection*, *description* and *matching*. Firstly a detector identifies local image regions (*keypoints*), which are both stable (consistently identified in all images) and accurate (well localised in all images). A descriptor, is then generated for each identified keypoint based on the pixels in its local neighbourhood. These descriptors must be distinct in order to differentiate between similar objects, yet robust to enable matching in the presence of noise, localisation error, and changes in scale, viewpoint and illumination. The state of the art comprises of numerous types and variations of local feature detectors and descriptors and there is none that is superior for all applications.

SURF (Speeded-Up Robust Features) descriptor is evaluated in our study as it is more computationally efficient than SIFT, while maintaining similar levels of matching performance [1]. The most used state of the art features (including SURF), do not utilise any colour information to generate their descriptors. They rely on the intensity pixel gradient information around the local neighbourhood of the keypoint. This generates highly unique descriptors, however colour information has the potential to provide further distinctiveness to a local descriptor to enhance its matching capabilities. This paper proposes a feature matching strategy that increases the number of identified correspondences by differentiating between ambiguous descriptors with colour information. Simple colour histograms are generated to encapsulate the colour of the local neighborhood around a candidate keypoint.

II. Background

Several comprehensive studies have shown that local distribution-based descriptors perform significantly better than other features [2]. Among them SIFT [2], has been proven to be the dominant feature in state of the art recognition systems due to its robustness (which in turn makes it computationally expensive). In [3], the authors evaluated different colour descriptors in a structured way, comparing their invariant properties to photometric transformations. Their results indicated that the use of colour descriptors indeed enhances object and scene recognition categorisation results (albeit by 8%). In their study, the OpponentSIFT descriptor came out as the most appropriate to use in general conditions. It comprises of the concatenation of three SIFT descriptors (and thus highly computationally expensive), each having been extracted from a channel of the opponent colour space (equation 1). They showed also that the normalised colour space (equation 2, where μ denotes the mean of the channel and σ the standard deviation), had the best photometric invariant properties and will also be compared in our study.

$$\begin{pmatrix} O1 \\ O2 \\ O3 \end{pmatrix} = \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{pmatrix} \tag{1}$$

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} (R-\mu_R)/\sigma_R \\ (G-\mu_G)/\sigma_G \\ (B-\mu_B)/\sigma_B \end{pmatrix} \tag{2}$$

Since the categorisation improvements reported were not significant, we propose to use simpler colour descriptors that encapsulates the colour information within a keypoint's neighborhood while minimising the computational costs. These descriptors will then enhance the distinctiveness of the local feature by combining with the SURF descriptors.

III. Descriptor Matching Strategy

Normalised Cross Correlation (*NCC*) is used as the clustering metric to compare corresponding keypoint SURF descriptors between two images. Every keypoint in each image is labeled with the first 5 nearest neighbours (1-5NN) of its corresponding image (our experiments indicated the 5NN classification approach was superior to 1NN, 3NN and 7NN). To ensure maximum accuracy, a keypoint pair is matched only if they are mutual first nearest neighbours (1NN). Colour information is used then, for each point that does not have a mutual 1NN. In this case, a keypoint I_i, from image I, has 5NN candidate keypoints (I'_{NN1-5}) from image I' that it can be matched with. The keypoint I_i , is considered for matching if it is itself one of the 5NN of either I'_{NN1-5}, see Figure 1. A colour

978-0-7695-4629-2/11 $26.00 © 2011 IEEE
DOI 10.1109/IMVIP.2011.31

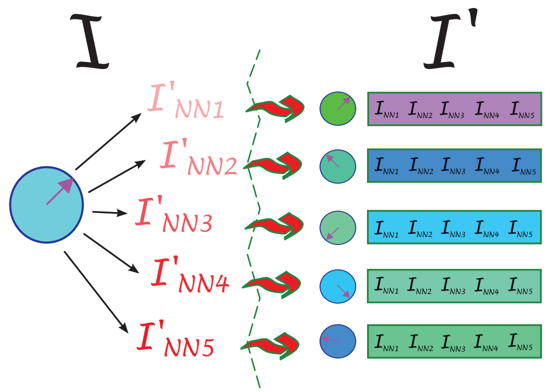

Fig. 1. Feature matching strategy based on mutual nearest neighbours between corresponding images I and I'.

histogram is generated for the neighbourhood around I_i and each of I'_{NN1-5} , and the keypoint that is more correlated to I_i (using NCC) is the chosen matched candidate.

IV. Colour Histograms

Five different colour spaces are used to compare the histograms: *RGB*, *N-RGB*, *HSV*, *N-HSV* and *OPPONENT*. The N-RGB space (normalised RGB) is obtained using equation 2, N-HSV is obtained by transforming the N-RGB into HSV (note that only the local neighbourhood is analysed during the normalisation). Finally, equation 1 transforms RGB into OPPONENT. A keypoint's colour descriptor is first composed of three separate histograms generated by quantising each colour channel of its local image neighbourhood into 16 colour bins. The frequency of each histogram bin is scaled as a percentage of the total pixel count of the keypoint's neighbourhood to be invariant to the size of the descriptor. The final descriptor of 48 dimensions is formed by concatenating the three colour channel histograms together.

V. Feature Matching Results

The experiments are carried out on a dataset of 27 image pairs, the corresponding images vary in illumination, viewpoint and translation changes. As a control, feature matching is performed on the images using only the SURF descriptors with the mutual 1NN NCC clustering strategy outlined in section III. These results are compared with the enhanced matching that uses colour descriptors from the colour spaces in section IV. All correspondances are filtered through a *RANSAC* algorithm that rejects any outlier pairs that do not fit in with the Fundamental matrix model. This model is not equally suitable to all image pairs, and can in cases reject correct and/or keep incorrect matches.

VI. Conclusion

The results in Figure 2 clearly show that utilising colour information increases the number of matches. N-HSV performs best on the illumination dataset with a mean improvement of 4.1%, HSV and OPPONENT are a close joint second (3.5%). For the translation/viewpoint datasets, HSV performs best with

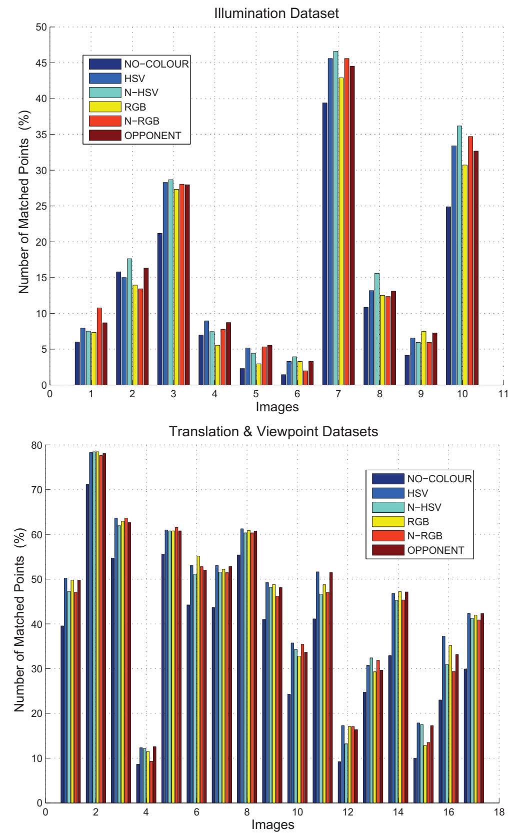

Fig. 2. Feature matching results for the three categories in the dataset. The results show the number of matched points as a percentage to the total keypoints extracted in an image.

a mean improvement of 8.97% , OPPONENT is again second (8.19%). The RGB colour spaces performed worst overall and are not recommended, as the colour information is correlated across all three channels. Each space's performance depends on the illumination conditions of an image, and the variation in the results suggests that the choice of colour space is not critical. What is more important to achieve better matching results in general conditions, is how the colour histogram themselves are generated. These simple histograms are not distinct enough and a more sophisticated way to encapsulate the colour information of a keypoint must be developed.

Acknowledgment

This research is funded by the Irish Research Council for Science, Engineering and Technology: funded by the NDP.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1615–1630, 2005.

[3] K. E. van de Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1582–1596, 2010.