Conditional Variance of Differences: A Robust Similarity Measure for Matching and Registration

Atsuto Maki and Riccardo Gherardi

Cambridge Research Laboratory, Toshiba Research Europe
Cambridge, United Kingdom

Abstract. This paper presents a new similarity measure, the *sum of conditional variance of differences* (SCVD), designed to be insensitive to highly non-linear intensity transformations such as the ones occurring in multi-modal image registration and tracking. It improves on another recently introduced statistical measure, the sum of conditional variances (SCV), which has been reported to outperform comparable information theoretic similarity measures such as mutual information (MI) and cross-cumulative residual entropy (CCRE). We also propose two additional extensions that further increase the robustness of SCV(D) by relaxing the quantisation process and making it symmetric. We demonstrate the benefits of SCVD and improvements on image matching and registration through experiments.

1 Introduction

A robust similarity measure between different regions of images plays a fundamental role in several image analysis applications such as stereo matching, motion estimation, registration and tracking. Similarity measures commonly used in these tasks, SSD or NCC for example, can at most cope with linear variations of intensity, such as global changes in gain and bias. Matching and registration techniques in general need to be robust to a wider range of transformations that can arise from non-linear illumination changes caused by anisotropic radiance distribution functions, occlusions or different acquisition processes [1] (e.g. visible light and infrared, those employed in medical imaging). These more challenging contexts, which represent the main focus of this article, have been extensively explored in the literature.

Most of the existing methods for computing similarity measures across multi-modal images are based on information theoretic approaches and make use of the probability of the intensity co-occurrence. The seminal works on mutual information (MI) [2,3] introduced the use of joint intensity distributions, recognising the statistical dependence between intensities of corresponding pixels. Other statistical dependencies have also been explored: cross-cumulative residual entropy (CCRE) [4] for example measures the entropy defined using cumulative distributions. The increased resilience to non-linear intensity transformations however comes at the cost of a higher computational complexity than conventional sum-comparing metrics, whose complexity is linear with respect to the number of elements.

G.L. Gimel' farb et al. (Eds.): SSPR & SPR 2012, LNCS 7626, pp. 657–665, 2012.

A recently proposed method, called the *sum of conditional variances* (SCV) [5], also uses the joint distribution of image intensities, but generates it directly as a histogram. Intuitively, SCV exploits a statistical property assuming that a group of pixels clustered by neighbouring intensities in the first image should be similarly clustered in the second, even if their mapped ranges are very different. SCV was originally developed in the context of medical image registration [6] and therefore aimed at being robust against non-linear intensity variations such as those occurring when capturing images through different acquisition modalities. It has been shown to have a larger convergence basin than MI's Parzen window approach in medical alignment tasks [7]. These results have been confirmed in the context of visual tracking [8], showing SCV to have better performance than several competing approaches in terms of convergence radius, computational complexity and stability (quantified by the number of iteration necessary for convergence). SCV is closely related to the correlation ratio [9], but has a lower computational complexity and is therefore more amenable to efficient optimization strategies.

In this paper, we introduce a new similarity measure called the *sum of conditional variance of differences* (SCVD). In the original SCV formulation, the reference image is used solely in its quantized form for generating a partition to be applied to the second image (i.e. the set of *conditions*). This process discards a significant amount of information. Assuming the intensity map to be weakly order preserving, whether directly or inversely, we show that the information loss can be mitigated employing the variance of *intensity differences*, leading to a more discriminative measure without increase in the computational complexity. We also generalise the computation of conditions, improving both our matching measure SCVD and the original SCV implementation.

The contribution of this paper is thus two-fold:

1. we introduce a novel similarity measure, the *sum of conditional variance of differences* (SCVD) and show its superior performance in comparison to other metrics designed against non-linear intensity variations,
2. we generalise the definitions of conditions, leading to improvements for both our formulation and the original SCV approach.

The rest of the paper is organized as follows: the next section will contain a brief description of the SCV algorithm, followed by the description of our proposal (SCVD) and its extensions respectively in section 3 and 4. We will evaluate the performance of the novel matching measure in section 5, focusing on matching and registration tasks. Finally, section 6 will report our conclusions.

2 The Sum of Conditional Variances

Given a pair of images X and Y, the sum of conditional variances (SCV) matching measure [5] prescribes to partition the pixels of Y into n_b disjoint bins $Y(j)$ with $j = 1, ..., n_b$, corresponding to bracketed intensity regions $X(j)$ of X (called

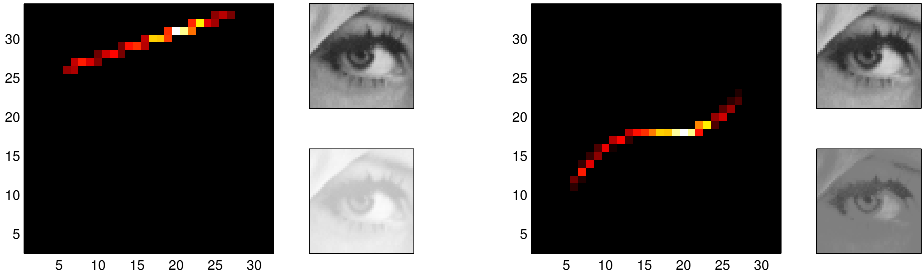


Fig. 1. Joint intensity histograms. A joint histogram H_{XY} can be interpreted as non-injective relation that maps the ranges of two images. **On the left:** the resulting joint histogram after linearly reducing the contrast of the reference image. **On the right:** the joint histogram for a non-linear intensity map. Hotter (brighter) colors correspond to more frequently occurring values.

the *reference* image). The value of the matching measure is then obtained summing the variances of the intensities within each bin $Y(j)$.

$$\mathcal{S}_{SCV}(X,Y) = \sum_{j=1}^{n_b} E[(Y_i - E(Y_i))^2 \mid X_i \in X(j)] \tag{1}$$

where X_i and Y_i with $i = 1, ..., N_p$ indicate the pixel intensities of X and Y respectively, N_p being the total number of pixels. The *conditions* that appear in the sum are obtained uniformly partitioning the intensity range of X.

The behaviour of SCV can be characterised by the joint histogram H_{XY} of X and Y. As shown in figure 1, the joint histogram can be interpreted as non-injective relation that maps the range of the first image to the second one. The set of pixels that contributed to the non zero entry of each column (row) corresponds to one of the regions selected by the j-th condition.

The number of discretisation levels n_b is problem specific; for images quantised at byte precision, a typical choice is usually $n_b = 32$ or 64 [8]. Larger intervals can help in achieving a wider convergence radius and offer more resilience to noise (the matching measure will not change as long as the pixels do not cross the current bin boundaries). On the other hand, narrow ranges will boost the matching accuracy and reduce the information that is lost during the quantisation step.

3 Sum of Conditional Variance of Differences

According to the SCV algorithm, the reference image is used solely to determine the subregions in which the variances of equation 1 should be computed. In this section, we present a new similarity measure based on the *conditional variance of differences*, which uses all the information present in both images leading to a more discriminative matching measure. We also propose two generalisations of the conditionals computation, which further increase the robustness of our approach.

3.1 Variance of Differences

We first define the *variance of differences* ($\mathcal{VD}$) as the second moment of the intensity differences between two templates:

$$\mathcal{VD}(X,Y) = Var[\{Y_i - X_i\}_{i=1\ldots N_p}] \tag{2}$$

The variance of differences is minimal when the distribution of differences is uniform. It is bias invariant, scale sensitive and proportional to the zero-mean sum of squared differences (sometimes called ZSSD or ASSSD in the literature). This last fact can be trivially verified from eq. 2:

$$\mathcal{VD}(X,Y) = E[(Y - X - E(Y - X))^2] \tag{3}$$

$$\propto \sum_i [(Y_i - E(Y_i)) - (X_i - E(X_i))]^2, \tag{4}$$

where the mean of an image is understood to indicate its element-wise mean.

3.2 Sum of Conditional Variance of Differences

Given two images X and Y, we define the *sum of the conditional variance of differences* (SCVD) as the sum of the variances over a partition of their difference. As before, the subsets are selected bracketing the range of the reference image to produce a set of bins $X(j)$. In order for the difference to be meaningful, the two signals should be in direct relation; since the matching measure need be insensitive to changes in scale and bias, we maximise direct relation by adjusting the sign of one of them in accordance with eq. 6. In symbols:

$$\mathcal{S}_{SCVD}(X,Y) = \sum_{j=1}^{n_b} \mathcal{VD}(X_i, \varPhi Y_i \mid X_i \in X(j)), \tag{5}$$

$$\varPhi = \varGamma\left(\sum_{j=2}^{n_b} \varGamma\left(E(Y_i \mid X_i \in X(j)) - E(Y_i \mid X_i \in X(j-1))\right)\right), \tag{6}$$

where $\varGamma$ indicates the step function mapping $\mathbb{R}$ to $\{-1, 1\}$. $\varPhi$ encodes a cumulative result of comparisons between a pair of $E(Y_i)$ in the adjacent histogram bins, so that the sign is properly adjusted. Hence, the requirement for the mapping from X and Y is to be weakly order preserving (the function should be monotonic but is not required to be injective). This restriction, not present in the original SCV formulation, makes it possible to make better use of the available information and largely valid, e.g. between signals captured for the same target with different modes.

4 Generalising the Conditions

Uniformly partitioning the intensity range of X into equally sized bins $X(j)$ can lead subpar performances when the intensity distribution is uneven: poorly sampled intensity ranges are noisy and their variance unreliable. Overly sampled regions of the spectrum conversely lead to compressing many pixels into a single bin,

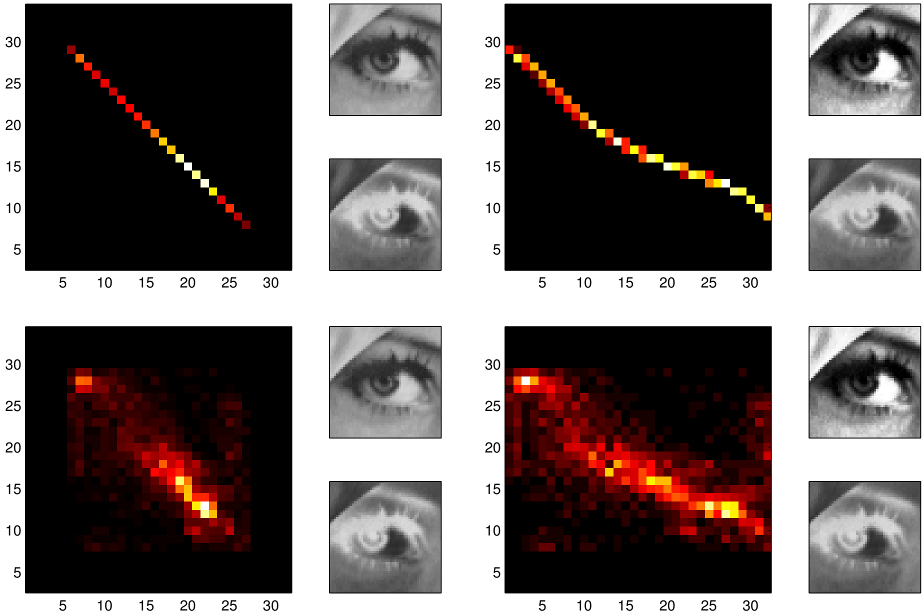


Fig. 2. Effects of quantisation and displacement. On the top row: H_{XY} for a pair of aligned images. Bottom row: H_{XY} for the same pair plus displacement. Left column: H_{XY} using uniform quantisation of intensity range. Right column: H_{XY} by using histogram equalised intensities for the reference image.

discarding a large amount of useful information in the process. The procedure is also inherently asymmetric, producing in general different results when swapping the images involved. In this section we discuss two non-mutually exclusive modifications of our proposal in order to deal with these issues. Each one of them provides an independent performance boost to the baseline approach described.

4.1 Uniform Quantizations

In fig. 2 (top-left) is shown the joint histogram between an image and its gray scale inverse. As it can be seen, the bins corresponding to the low and high end of the intensity spectrum are not receiving any vote, thus compressing the image information into a smaller number of regions.

To achieve a uniform bin utilisation, we perform histogram equalisation on the reference image X. Figure 2 (right) shows an H_{XY} generated by replacing the input reference image X with its histogram equalized version, achieving full utilisation of the entire dynamic range.

On the bottom row of fig. 2 are shown the original and histogram equalised version after applying a 5 pixel displacement to one of the images. As a result, the entries are more scattered and less sharp. As in the previous case, the non equalised version does not make full use of the available bins; the equalized one, shown at the bottom right spreads the vote over a larger area, affecting the variance computation and resulting in a more discriminative measure.

4.2 Bi-directional Quantisations

Both SCV and SCVD are structurally asymmetrical since only one of the images is used to define the partitions in which to compute the variance. Generally, $\mathcal{S}_{\{SCV,SCVD\}}(X,Y) \neq \mathcal{S}_{\{SCV,SCVD\}}(Y,X)$ because the two quantities are computed over different subregions which depends on the reference image. As far as the task of image matching is concerned, no particular reason exists in choosing one image over the other as the reference; the process of quantization can thus be symmetrised computing $\mathcal{S}_{\{SCV,SCVD\}}$ bi-directionally:

$$\mathcal{S}^{B}_{\{SCV,SCVD\}} = (\mathcal{S}_{\{SCV,SCVD\}}(X,Y) + \mathcal{S}_{\{SCV,SCVD\}}(Y,X)) \ / \ 2 \ . \qquad (7)$$

Given the characteristics of SCVD (SCV), in presence of uneven quantizations one direction is usually much more discriminative than the other. The above formula is capable of successfully disambiguating such situations.

5 Experimental Evaluation

Experiment I. In order to compare our proposal, its variations and the original SCV approach, in this first experiment we study the discriminativeness of each one of them for increasing, isotropic displacements. We selected an image location, a direction and a displacement all at random, computing the measure between the selected reference window and the template after applying the translation. Notice that the template is negated in order to simulate multi-modal inputs. The size of the region was fixed to 50×50 pixels while the maximum distance was set to be half of its edge length, i.e. 25 pixels. The results are shown for a single image (the *peppers* image included in Matlab) but the plot of figure 3 is similar for any non-periodic, non-uniform picture.

Figure 3 was produced averaging 20,000 iterations of this procedure, to remove the effects of noise (each single trial is roughly monotonic). As it can be seen, all $\mathcal{S}_{SCVD}$ versions are better at discriminating the minimum. Histogram equalized and symmetric variants obtain steeper gradients for both SCV and SCVD. When utilising both improvements, SCVD shows a nearly constant slope, a crucial property in order to use optimization algorithms based on implicit derivatives.

Experiment II. We now compare the performance of different similarity measures on a synthetic registration task using a gradient descent search; given a random location and displacement as before, we optimize the cost function following the direction of the steepest gradient. The procedure terminates when reaching a local minima or the maximum number of allowed iterations (set to 50 in our experiments). Figure 4 was obtained averaging 4000 different trials; as it can be seen, each SCVD version beats the equivalent SCV measure using the same set of variants, which provide a non negligible performance boost.

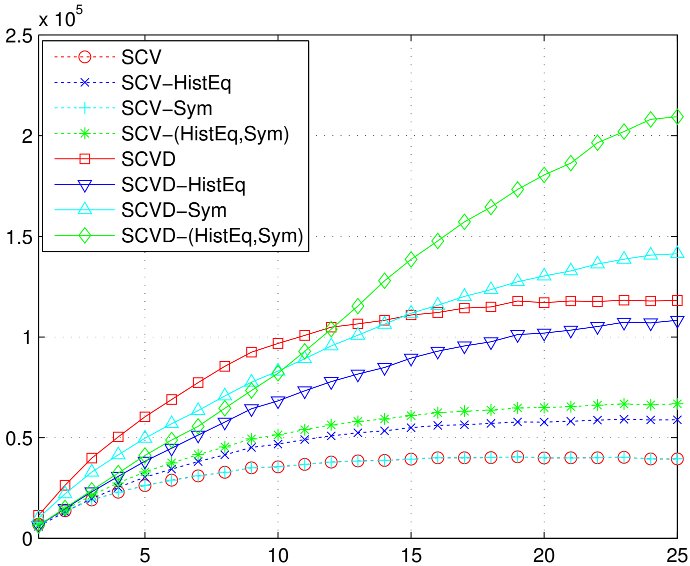


Fig. 3. Matching measure vs. displacement. We compare our proposal, its variations and the original SCV approach over random displacements within an image. SCVD plus both extensions results in the most discriminative measure, with a nearly constant slope across the entire search domain.

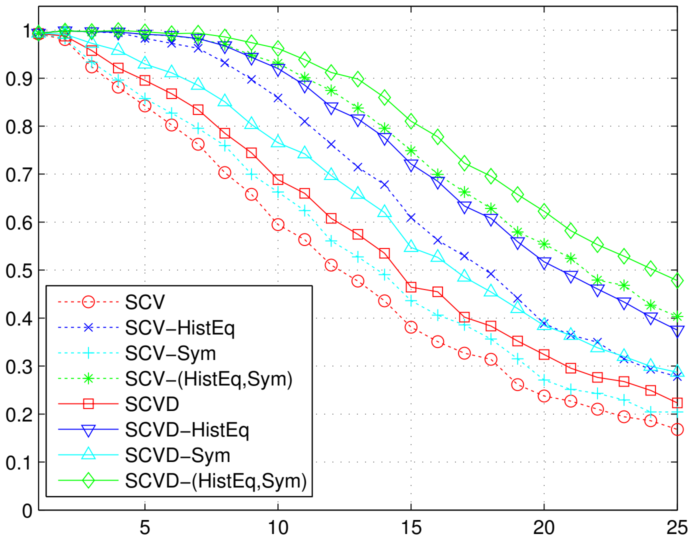
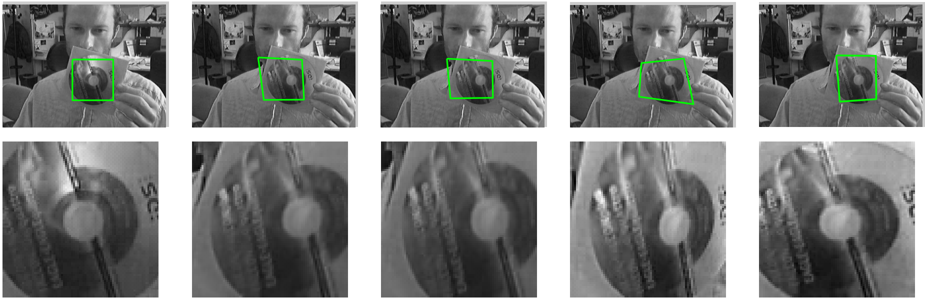


Fig. 4. Convergence vs. displacement. The plots show the convergence rate as a function of the distance between the reference and displaced window. We compare our proposal, its variations and the original SCV approach over random displacements within an image.



(a) (b) (c) (d) (e)

Fig. 5. Registration experiment. (a) Input frame with reference region marked green. (b-e) Registrations by MI, CCRE, SCV and SCVD. On the second row are shown the registered regions backwarped to the template (sequence part of the ESM project, http://esm.gforge.inria.fr).

Experiment III. In our final experiment we compare the performance of several similarity measures on a tracking task over a real image sequence. Figure 5 (a) shows one of the frames of the sequence, and its reference template. The subsequent frame has both photometric and geometric deformations; in figure 5 (b-e) we display the registration results respectively for MI, CCRE, SCV and SCVD, showing both the best matching quadrilateral on the frame and the regions backwarped to the reference. The results with SCVD and SCV are by our implementation while those with MI and CCRE are by an implementation by [8] on the basis of the software presented in [10].

6 Conclusions

We presented a new statistical similarity measure, the *sum of the conditional variance of difference* (SCVD), tailored for robustly matching two image regions in presence of non-linear intensity transformations. Under the assumption of the transfer function being weakly order preserving, we have shown our proposal to outperform the *sum of the conditional variance* (SCV), a recent algorithm that was already shown to be competitive with the current state of the art. We also developed two non mutually exclusive improvements that can make both SCV and SCVD more discriminative at a negligible computational cost. Although we have demonstrated the benefit of SCVD in the context of image matching and registration, its principle is applicable to measure the similarity of two 3D volumes.

References

1. Irani, M., Anandan, P.: Robust multi-sensor image alignment. In: ICCV, pp. 959–966 (1998)

2. Viola, P.A., Wells III, W.M.: Alignment by maximization of mutual information. International Journal of Computer Vision 24, 137–154 (1997)
3. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16, 187–198 (1997)
4. Wang, F., Vemuri, B.C.: Non-rigid multi-modal image registration using cross-cumulative residual entropy. International Journal of Computer Vision 74, 201–215 (2007)
5. Pickering, M.R., Muhit, A.A., Scarvell, J.M., Smith, P.N.: A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In: Int. Conf. of IEEE Engineering in Medicine and Biology Society, pp. 5821–5824 (2009)
6. Zitová, B., Flusser, J.: Image registration methods: a survey. Image Vision Comput. 21, 977–1000 (2003)
7. Pickering, M.R.: A new similarity measure for multi-modal image registration. In: International Conference of Image Processing, pp. 2273–2276 (2011)
8. Richa, R., Sznitman, R., Taylor, R.H., Hager, G.D.: Visual tracking using the sum of conditional variance. In: IROS, pp. 2953–2958 (2011)
9. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 1115–1124. Springer, Heidelberg (1998)
10. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. International Journal of Computer Vision 56, 221–255 (2004)