Feature Regression for Multimodal Image Analysis

Michael Ying Yang, Xuanzi Yong, Bodo Rosenhahn
Institute for Information Processing (TNT), Leibniz University Hannover
Appelstr. 9A, 30167 Hannover, Germany
{yang,yong,rosenhahn}@tnt.uni-hannover.de

Abstract

In this paper, we analyze the relationship between the corresponding descriptors computed from multimodal images with focus on visual and infrared images. First the descriptors are regressed by means of linear regression as well as Gaussian process. We apply different covariance functions and inference methods for Gaussian process. Then the descriptors detected from visual images are mapped to infrared images through the regression results. Predictions are assessed in two ways: the statistics of absolute error between true values and actual values, and the precision score of matching the predicted descriptors to the original infrared descriptors. Experimental results show that regression methods achieve a well-assessed relationship between corresponding descriptors from multiple modalities.

1. Introduction

In recent years multisensor data fusion receives more and more attention [10]. The integration of images from multiple modalities can provide complementary information and therefore the accuracy increases with an observed and characterized quantity. In the domain of computer vision, the detected points can be represented by some descriptors. In this way a point with its surroundings is described by a vector. Therefore, we represent an image with a set of vectors, which get rid of the noises and some unnecessary information. Moreover the computational costs are also reduced as well as the memory costs, which shows to be more efficient.

Given multi-modal images, a series of applications are provided such as matching objects and scene registration. It is easy to obtain the interest points from the given images and reform them by some feature descriptors. However, it comes to a question if there is some relationship between the two corresponding feature vectors. Or can we get the feature descriptor of a point in infrared image by the corresponding point in visual image. Moreover, with a mapping function, a descriptor is mapped to an infrared image as a new vector. So how would this new vector looks like, where might this new vector locate in the infrared images. In this paper, we address the aforementioned problems using regression methods. To the best of our knowledge, there is no existed work on analyzing the relationship among descriptors in multimodal images.

In this work, we analyze the behavior of features in multimodal images with focus on visual and infrared images. In order to get a reliable result, we construct the datasets with different types of images from different categories. We present regression results of feature descriptors from visual images to infrared modality, which indicates the existence of the relations between the descriptors. The regression is worked in two ways: linear regression and Gaussian process for regression (GPR). The former has a computational advantage that it runs faster and costs lower than other common regression methods. And the latter can obliquely represent the underlying regression function without claiming, but rigorously. As a result, the descriptors of points detected in visual images are mapped as the descriptors from infrared images. We evaluate the performances of linear regression mainly by the value of coefficient of determination and the results are evaluated by the mean and variance of error between the descriptor vectors. In order to assess the performance of Gaussian process regression, we apply the regression result to the application of matching. The results are evaluated by the precision of matching. Moreover the regression error is considered as the criterion as well. From the results, we can find that, based on specific covariance functions and inference methods, the regression process performs well that the predicted descriptors are similar to the actual descriptor vectors. Example results of GPR are shown in Fig. 1.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 presents our two regression methods for multimodal image analysis: linear regression and Gaussian process for regression. In Section 4, we provide a short review of related descriptors. Section 5 shows the results of two regression methods for the feature descriptors in multimodal images. Finally, this work is concluded in Section 6.

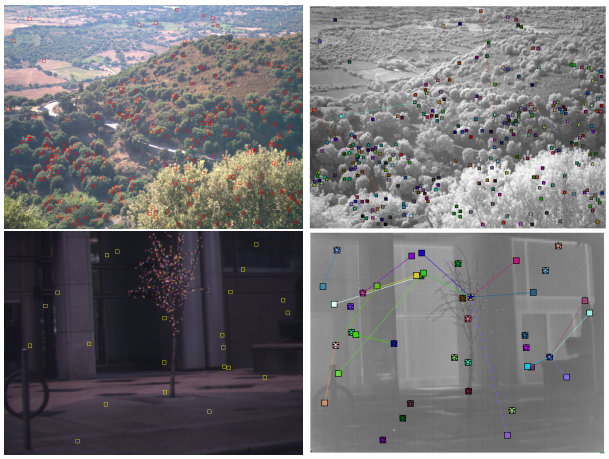

Figure 1. Example results of GPR using SIFT. The left column are original visual images with detected points and on the right side are infrared images with relocated descriptors that are achieved by Gaussian process regression.

2. Related Work

On account of the development of sensor fusion technique, many applications were developed under multiple modality. [13] explored a statistic research on analyzing the significant characters on infrared images. Compared to corresponding visual images, the infrared images had noticeably less texture indoors because of the homogeneous temperature. Further, the joint wavelet statistics presented strong correlation between object boundaries in visual and infrared images, which could be used in vision applications with the combined statistical model. Moreover an overview of registering different types of sensors was provided by [17]. [8] studied an approach to multimodal image registration based on corners and Hausdorff distance. The approaches using the mutual information as the matching criterion are the state-of-the-art technique in multispectral matching [12, 9]. Due to the points and the contours of infrared images are different enough relative to visual images of the same scene, this region-based technique performs relatively well. [7] implemented a line-based global transformation using the edge properties for the image registration between visual images and infrared images. Besides, [6] provided a feature based matching and multimodal RGB to NIR registration with multispectral interest points. Furthermore an experiment for multimodal 2D and 3D face recognition was presented by [4]. [1] pursued on the problem of matching images with disparate appearance arising from factors such as dramatic illumination (day vs. night), time period (historic vs. new) and rendering style differences. By using the eigen-spectrum of the joint image graph, the persistent features were detected and matched into pairs.

3. Descriptor Regression

Given sets of descriptors $\mathbf{DA}$ detected from visual images and $\mathbf{DB}$ from infrared images, regression analysis is to estimate the relationships between $\mathbf{DA}$ and $\mathbf{DB}$. By means of the relationship, which might be implicit or explicit, the predictions $\mathbf{DA}'$ come out that the descriptor is mapped from visual image to infrared modality. The regression process includes two techniques for modeling and analyzing variables. Since linear regression is a common and light method used in stochastic problems, it is considered in this work at first. Then Gaussian process is used as an advanced method, which can especially solve non-linear problems.

3.1. Linear Regression

In statistics, linear regression is an approach to model the relationship between a scalar dependent variable and one or more explanatory variables, in which data are modeled by linear functions and unknown model parameters are estimated from the data [3].

Upon to the principle of linear regression, a descriptor Da with n dimensions from visual image is mapped to a descriptor as Db in the same dimension in infrared image through a matrix as linear transformation, which is given as:

$$Db = Da \times H \tag{1}$$

where H is a $n \times n$ matrix. H is calculated by training and then used to predict the new input Da forward to a Db. The process of linear regression is implemented by the method of Least squares. It is a technique for mathematical optimizing that the sum of the squares of the errors is minimized by equating its gradient to zero and then the regressors are obtained through the mean value.

3.1.1 Regression estimating

The regression procedure is estimated by a set of statistics, such as R^2, F, p and the estimate of the error variance $err.var$. R^2 is the *coefficient of determination* defined as

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}, \tag{2}$$

and SS_{tot} is the total sum of squared errors in the model that does not use the independent variable, and SS_{res} is the sum of squared errors in the linear model. It is a very important indicator to state if the regression is efficient while it informs the *goodness of fit* of a model. In regression, R^2 represents the percent of the data that is the closest to the line of best fit, in other words, it informs how well the regression line approximates the real data points. The F statistic is the test statistic of the F-test on the regression model, for

a significant linear regression relationship between the response variable and the predictor variables. P-value p is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. When the p-value is less than the given significant level, in usual case as 0.05, the null hypothesis will be rejected. By using these arguments, the performance of linear regression is evaluated.

3.1.2 Predictions evaluating

After the regression procedure, the linear transformation matrix H is obtained and then used to predict the new descriptor Da forward to a Da'. For sake of predictions evaluating, we compare the prediction Da' and the true descriptor vector Db by the absolute difference between corresponding components in vectors as $\varepsilon = |Da' - Db|$. Two parameters are used to evaluate, that is $mean$, which is the average value of ε and the variance of ε.

However, the method of linear regression can not solve non-linear problems. Hence we use Gaussian process for regression as an advanced method, in which a specific model need not to be claimed at first.

3.2. Gaussian Process for Regression

Given some noisy observations of a dependent variable, the estimate of a new value x comes out easily by using a function $f(x)$, which can describe the distribution of the observations. Rather than a specific model which the claimed function $f(x)$ relates to, a Gaussian process can represent $f(x)$ obliquely, but rigorously [15]. That is so-called Gaussian Process Regression (GPR).

Taking account of the noise on the observed target values from measurement errors and so on, which are given by

$$t_n = y_n + \epsilon_n \tag{3}$$

where $y_n = f(x_n)$, and ϵ_n is a random noise variable whose value is chosen independently for each observation n.

The conditional distribution of t_{N+1} given target values $\mathbf{t} = (t_1, \ldots, t_N)^{\mathrm{T}}$ is itself Gaussian-distributed as the form:

$$t_{N+1}|\mathbf{t} \sim \mathcal{N}(\mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{t}, c - \mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{k}). \tag{4}$$

The mean, $\mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{t}$, is known as the *matrix of regression coefficients*, and the variance, $c - \mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{k}$, is the *Schur complement* of $\mathbf{C}_N$ in $\mathbf{C}_{N+1}$. These are the key results that define Gaussian process regression. While the vector $\mathbf{k}$ is a function with respect to the test input value $\mathbf{x}_{N+1}$, the predictive distribution is a Gaussian depended on $\mathbf{x}_{N+1}$.

As a crucial component of a Gaussian process predictor, covariance function controls how much the data are smoothed in estimating the unknown function [15]. Two functions are considered: the *squared exponential* (SE) covariance function has the form

$$k_{SE}(r) = exp(-\frac{r^2}{2\ell^2}), \tag{5}$$

with parameter ℓ defined as *characteristic length-scale*. This covariance function has sample functions with infinitely many derivatives and thus is very smooth. Another is rational quadratic (RQ) covariance function

$$k_{RQ}(r) = (1 + \frac{r^2}{2\alpha\ell^2})^{-\alpha} \tag{6}$$

with $\alpha, \ell > 0$, which can be regarded as a *scale mixture* (an infinite sum) of squared exponential (SE) covariance functions with different characteristic length-scales (sums of covariance functions are also a valid covariance).

The descriptors are treated in two ways:

3.2.1 Global descriptors

Assuming a particular structure, where the covariance function is set as the *squared exponential* function and the mean of Gaussian process is defined as zero like the assumptions in most cases. In addition, the Expectation propagation (EP) is applied as the inference function and the likelihood function is in the form of Laplace. The parameters of covariance function are initialized with zero at first and later they are optimized by minimizing their negative log marginal likelihood.

3.2.2 Local descriptors

The goal of this part is to check the potential location relationship of the descriptors. Based on the training data, the descriptor vectors $\mathbf{DA}$ from visual images are mapped to the infrared images as $\mathbf{DA}'$. For each descriptor Da in the set $\mathbf{DA}$, we are looking for the most similar descriptors among all the descriptors $\mathbf{DB}$ in infrared images. In other words, a vector is predicted by a descriptor from visual image. And the task is to check the location of this vector in the infrared image.

The processing procedure is as following: first the interest points are detected from infrared and visual images and then represented by feature descriptors. Hence we obtain a set of vector pairs. Each vector consists of two parts, the descriptor of the interest points and its location. The next step is to obtain the predictions by Gaussian process. The initial hyperparameter of covariance function is set with 0.7. And then for one prediction vector, find the closest vector among all the original descriptor vectors in infrared images by using the Euclidean distance between two vectors. Moreover min-pooling approach is used to avoid too many incorrect

matchings. In practice, assuming the infrared image and visual image display completely the same scene. Five candidates are chosen with most similar vectors. And then the prediction is determined to locate in the position of its nearest candidate.

4. Feature Descriptors

Descriptors are used to represent the image structure in spatial neighborhoods at a set of feature points. There are various kinds of descriptors, and we can choose an appropriate one based on the application. In this section, we will present four descriptors used in this work, that is SIFT, SURF, LBP and HOG.

4.1. Scale-invariant Feature Transform (SIFT)

SIFT (Scale-invariant feature transform) is based on the interest points detected by Difference of Gaussian [11]. The descriptor records the direction for each interest point, thus it has good scale and rotational invariance. A key point is characterized with location, scale and direction. The orientations of 16×16 neighbors of each keypoint are calculated and then projected into one of eight directions with 4×4 region. Subsequently, a histogram is built with 8 bins, which indicate 8 directions. As a result, the descriptor is in the form of vector with 128 dimensions. With the help of this descriptor, we can match key points between images.

4.2. Speeded Up Robust Feature (SURF)

Speeded Up Robust Feature (SURF) is an improvement of SIFT, which is first presented by [2]. It is claimed that it performs excellent on repeatability, distinctiveness and robustness. The interest points are detected using Hessian matrix, that is named as Fast Hessian detector, which is calculated for each point. To solve it, SURF makes efficient use of integral images. Then by comparing each point with its 26 neighbors on the same octave and the octave above and below, the points with maximum or minimum responses are considered as interest points after filtering by given threshold. The descriptor is based on sum of Haar wavelet responses within the region in the size of 4×4, instead of histogram in SIFT, which is in the form as:

$$\sum d_x \quad \sum d_y \quad \sum |d_x| \quad \sum |d_y|,$$

d_x and d_y are the filter responses to the Haar wavelets. Thus the output of SURF is a feature vector with 64 dimensions.

4.3. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a type of texture spectrum model proposed in [16] and first described by [14]. In this approach, an examined window is first divided into 16×16 cells. And then for each pixel in a cell, comparing the gray-value with other eight neighbors. It is assigned as 1, when the neighbor is greater than center pixel. Thus, an 8 bit binary pattern comes, i.e LBP. Compute the histogram of the frequency of each binary number occurring over the cell and normalize. The feature vector for the window should be the concatenate normalized histograms of all cells.

4.4. Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) is first represented by [5], which focuses on pedestrian detection at that time. And the essential idea behind the Histogram of Oriented Gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. To implement it, the image need to be divided into small connected regions, called cells. And then compute the gradient for each pixel in the region of a cell. The histogram of gradient in each cell is the descriptor for the cell and the combination of these histograms present the descriptor. In some advanced process, the cells are grouped into larger spatial blocks and these blocks are normalized separately. As a result, the final descriptor is exact the vector composed of all the components of the normalized cells by the blocks in the detection window.

5. Experiments

We construct the experiments to regress the descriptors by using linear regression and Gaussian process. And the results are assessed by the criteria of $error$ and the precision of matching.

5.1. Datasets

Three datasets are used in this paper, namely RGB-NIR scene dataset from EPFL-IC-IVRG[1], OutdoorUrban by DGP-UofT[2] and MoCap from LUH-TNT[3].

These three datasets contain different image types with different viewpoints. The dataset RGB-NIR consists of numbers of images captured in RGB and Near-infrared (NIR) by visible and NIR filters using separate exposures from modified SLR cameras. There are totally 9 categories such as field, forest, mountain and water. And the images in the dataset OutdoorUrban are fully about the views around the city, such as cars, buildings and some other city scenes. Compared to the natural scenery, there are greater thermal variations in urban environments [13]. In this dataset, there are total 290 outdoor urban daytime image pairs as well as about 30 urban night-time image pairs. The data are captured by a single axis, multiparameter camera which combines an infrared camera and a visible light camera. And the content of dataset MoCap is the human motions such as

[1]Image and Visual Representation Group at EPFL
[2]Dynamic Graphics Project at University of Toronto
[3]Institute for Information Processing at Leibniz University Hannover

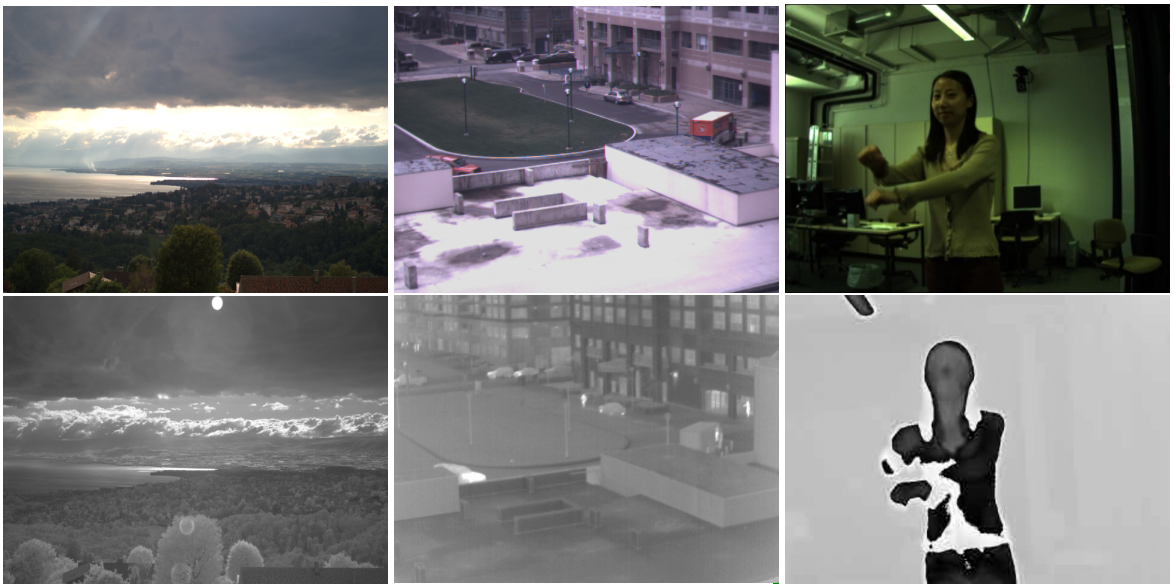

Figure 2. Sample images from the datasets RGB-NIR, OutdoorUrban and MoCap respectively. The images in the first row are visual RGB images and the images in the second row are the corresponding infrared (or near-infrared) images with respect to the images above.

Table 1. Information of datasets

Dataset	Type	Data number	Image size	Contents
RGB-NIR	near-infrared	370	640×480	nature views
OutdoorUrban	infrared	330	384×288	city views
MoCap	infrared	1300	640×480	human motions

Table 4. The test result of linear regression on HOG.

HOG	size	mean	var
RGB-NIR	70	0.0356	0.1111
Urban	27	0.2710	0.0656
MoCap	500	0.0243	5.1046e-04

Table 5. The test result of linear regression on LBP.

LBP	size	mean	var
RGB-NIR	70	449.8296	5.1483e+06
Urban	27	499.2019	4.8877e+06
MoCap	500	308.3616	2.0000e+06

waving, boxing and jogging indoors. Since the two cameras are set with a small baseline, the taken images are not identical in view, but nor too far away from each other. Some samples of the datasets are shown in Fig. 2 and the basic information is summarized in Table 1.

5.2. Results of Linear Regression

Considering the regression, 300 images in RGB-NIR are selected in the training set and the rest 37 images are kept for testing. For dataset OutdoorUrban, the size of training data is 100 and it is 27 of testing data. In dataset MoCap, there are 800 images in the training dataset and 500 images are regarded as testing data.

The regression is assessed by the results in Table 2 and Table 3. In the tables, the value of R^2 is around 0.9, which means that the regression function is much closer to the true values, and it understands the information of the data very well. Also most of the p-value in two tables are greater less than 0.05, so the null hypothesis is rejected, namely the linear model is correct for the data. But HOG in dataset OutdoorUrban with the value 0.0670 is an exception. In a word, the two descriptors are both regressed well with the training data, and they draw linear lines perfectly fitting to the points.

For testing, the $mean$ and var in Table 4 and Table 5 refer to the average value and variance of the error between the actual value and the true value. Since the components of HOG descriptor vectors are in the range of 0 to 1, that the mean error is only about 0.03 on RGB-NIR and MoCap indicates an excellent result. Meanwhile, the error of LBP seems much greater, but if they were normalized in the interval from 0 to 1, the average value of error is 0.0026 for example of RGB-NIR.

5.3. Results of GPR

5.3.1 GPR for HOG and LBP

By using these hyperparameters, the new feature descriptors are predicted. First set with 10 test data, the two criteria, $mean$ and $variance$ are computed as shown in Table 6. On datasets RGB-NIR and MoCap, the averages of absolute error are both under 0.05. Meanwhile the variances on the two dataset are in a great level as well.

Further, for the sake of analyzing the effect on training and testing dataset, we enlarge the size of training data to 100 images and the size of testing data is amplified to 50

Table 2. The statistics of linear regression on HOG.

HOG	size	R^2	F	p	err.var
RGB-NIR	300	0.9206	35.1212	0	0.0011
Urban	100	0.9072	2.5281	0.0670	0.0055
MoCap	1300	0.8482	101.2218	1.9102e-276	2.4714e-04

Table 3. The statistics of linear regression on LBP.

LBP	size	R^2	F	p	err.var
RGB-NIR	300	0.9735	8.3986	6.8146e-06	5.5635e+05
Urban	100	1	NaN	NaN	NaN
MoCap	1300	0.8842	49.4688	5.4552e-16	1.0459e+06

Table 6. The result of GP regression for HOG.

HOG	RGB-NIR	OutdoorUrban	MoCap
mean	0.0342	0.1376	0.0146
variance	5.8757e-04	0.0070	8.3017e-06

Table 7. The result of GP regression for HOG with 100 training and 10 testing data.

HOG	RGB-NIR	OutdoorUrban	MoCap
mean	0.0310	0.1416	0.0042
variance	4.0629e-4	0.0036	4.5617e-06

Table 8. The result of GP regression for HOG with 100 training and 50 testing data.

HOG	RGB-NIR	OutdoorUrban	MoCap
mean_err	0.0238	0.0280	0.0502
var_err	0.0004	0.0010	0.0024
fron_actual	21.1325	21.2427	20.8977
fron_true	21.4242	21.4240	21.2131

Table 9. The result of GP regression with exact inference method and Gaussian likelihood function.

HOG	RGB-NIR	OutdoorUrban	MoCap
mean_err	0.0251	0.0287	0.0500
corr2	0.8909	0.9609	0.8983
frob_actual	21.1551	21.2685	20.8614
frob_true	21.4242	21.4240	21.2131

respectively. Comparison the data in Tables 6, 7 and 8 in vertical direction, the result indicates that the size of neither training data nor testing dataset can effect the performance of Gaussian process heavily. Therefore, Gaussian process is robust and efficient with fewer training data.

Applying exact inference method and Gaussian as likelihood function, the sizes of training data and testing data are set with 100 and 50 respectively. Based on this setting, the process runs much faster than using EP inference method. From Table 9, we can see that the performance of evaluation is excellent, the average value of error is less than 0.03. According to Frobenius norm, the ratio between the actual value and true value is over 99%, which shows the similarity completely.

Also we consider the role of the initial value of the hyperparameters of the covariance functions. The values are set to change with step of 0.1. Since the procedure of parameter optimization is applied, the initial values make no sense to effect the result and performance.

For the purpose of LBP , the same process contents have been executed as HOG. However, the result turns that we can not obtain an answer to the regression for LBP.

5.3.2 GPR for SIFT and SURF

We consider squared exponential function (SE) and rational quadratic function as the covariance function for Gaussian process regression. And also the two inference methods: EP inference method and exact inference method, are applied in this work. In addition the mean value of Gaussian process is set as zero and 100. Thus, based on these three conditions, where each has two values, there are totally 2^3 combinations.

From the results of SIFT, the process with RQ covariance function by exact inference method outperforms than the others with an optimal result, especially in RGB-NIR with a precision over 90%. And the precisions on other sets are also acceptable with about 50%. But the SE covariance function is not fitted in this model. In addition, we can find that the value of mean has little effect on the results. On the aspect of SURF, both RQ and SE covariance functions perform well in this work. Based on the value of precision as well as the illustration of the example result images in Fig. 3, the regression results from EP method are totally failed in each situation despite of higher computational cost. In a word, the best performance is the process by exact inference method with RQ covariance function.

5.4. Linear regression vs. Gaussian process regression

For descriptor HOG, both linear regression and Gaussian process can predict reasonable mapping models from visual images to infrared images. Comparing the two methods,

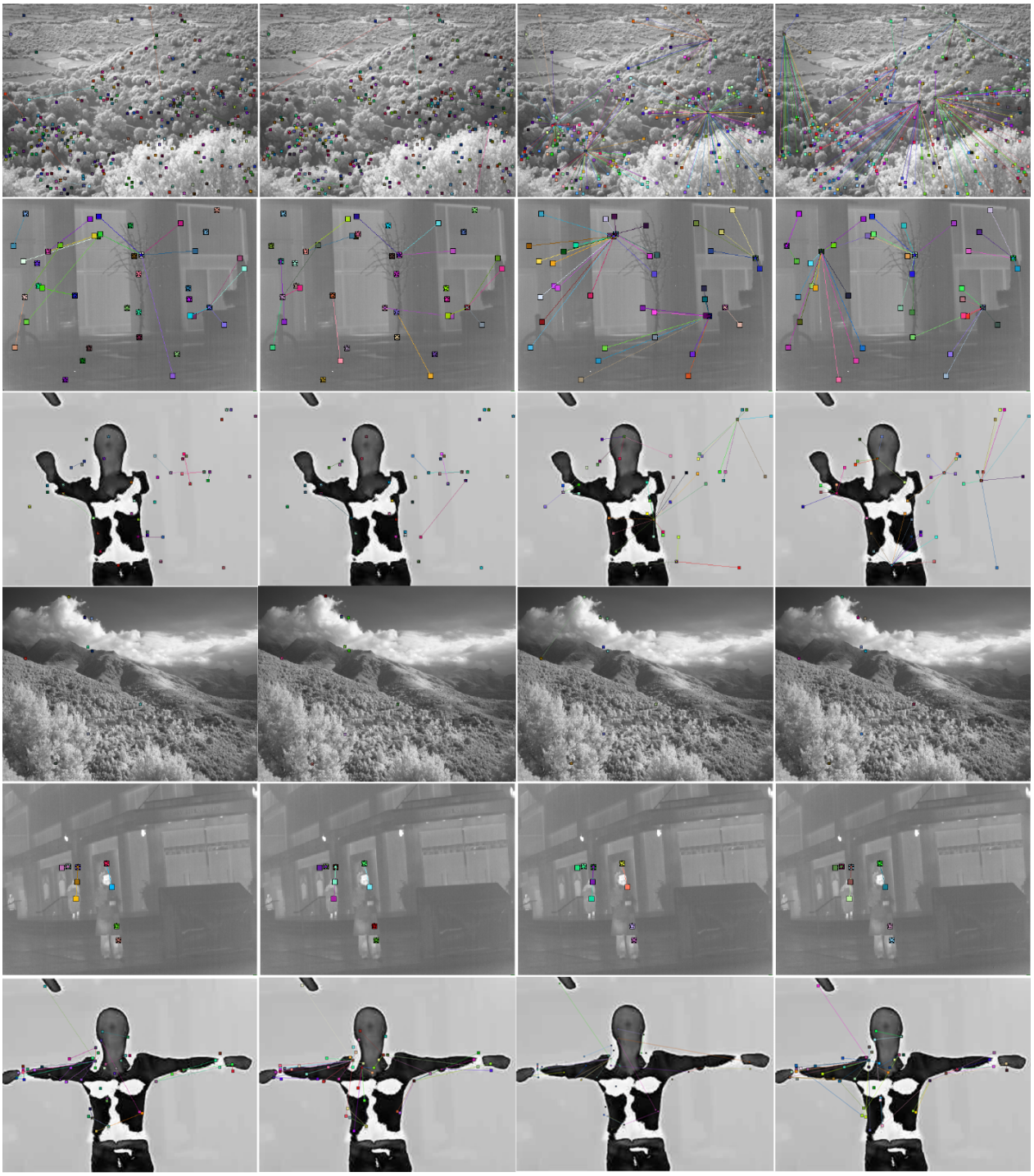

Figure 3. Example of descriptor matching. The descriptors are regressed by Gaussian process with exact inference method. First two columns refer to the results with RQ covariance function with mean value zero and 100, and the last two columns show the results with SE covariance function with mean value zero and 100. The squares refer to the detected points in visual images and the stars refer to the relocated descriptors in infrared images.

the results are shown in Table 10 depending on the condition of $error$ introduced before. Both approaches perform well with a low error. And it is obvious that Gaussian process performs better than linear regression, where $error$ by Gaussian process is extraordinarily small. Another advantage of Gaussian process appears when the training set is

Table 10. Comparison of linear regression and Gaussian process for HOG.

mean_error	Linear Regression	Gaussian Process
NIR	0.1074	0.0553
OutdoorUrban	0.7019	0.0505
MoCap	0.0243	0.0475

small. It means that in practice, the requisite prior knowledge is much less for GP than linear regression. However, for this instance, linear regression also performs well, so it is a good choice as well because the complexity and cost of linear regression is much lower than GP. Notice that actually for the dataset OutdoorUrban, it does not fit into a linear model. But Gaussian process can deal with linear model and also non-linear model problems. We can see that comparing to the error in OutdoorUrban by linear regression, Gaussian process is much better than it in this case.

6. Conclusion

In this paper, we have focused on the relationships among descriptors from infrared and visual image pairs. Three extensive datasets of infrared and visual image pairs are considered to explore the regressions. Between corresponding HOG and LBP, linear relations have been provided by least squares method with good regression qualities. This indicated the possibility to map a descriptor from visual image to infrared modality by a linear transformation. Furthermore, we have used Gaussian process for regression on HOG and LBP. The optimal regression results have been shown with small error by using squared exponential covariance function. The GPR results of SIFT and SURF have been evaluated by the application of matching. The process of SIFT with rational quadratic function as covariance function has a good performance by evaluating the precision score of matching. The results have presented not only the relationships of SIFT and SURF corresponding descriptors, but also the possibility of obtaining the relationship of descriptors in multi-modal images by means of Gaussian process. In addition, comparing the results of linear regression and GPR of HOG, Gaussian process performs better than linear regression but with a higher computational costs. For the future work, we will perform some regression analysis for other multimodal data, such as visual and depth images.

Acknowledgments

The work is funded by the ERC-Starting Grant (DYNAMIC MINVIP). The authors gratefully acknowledge the support.

References

[1] M. Bansal and K. Daniilidis. Joint spectral correspondence for disparate image matching. In *CVPR*, pages 2802–2809, 2013. 2

[2] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *ECCV (1)*, pages 404–417, 2006. 4

[3] N. H. Bingham, N. Bingham, and J. M. Fry. *Regression: Linear models in statistics*. Springer, 2010. 2

[4] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2d+ 3d face biometrics. *PAMI*, 27(4):619–624, 2005. 2

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005. 4

[6] D. Firmenichy, M. Brown, and S. Susstrunk. Multispectral interest points for rgb-nir image registration. In *ICIP*, pages 181–184, 2011. 2

[7] J. Han, E. J. Pauwels, and P. M. de Zeeuw. Visible and infrared image registration in man-made environments employing hybrid visual features. *Pattern Recognition Letters*, 34(1):42–51, 2013. 2

[8] T. Hrkać, Z. Kalafatić, and J. Krapac. Infrared-visual image registration based on corners and hausdorff distance. In *Scandinavian Conference on Image Analysis*, pages 383–392, 2007. 2

[9] J. P. Kern and M. S. Pattichis. Robust multispectral image registration using mutual-information models. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1494–1505, 2007. 2

[10] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28–44, 2013. 1

[11] D. Lowe. Object recognition from local scale-invari ant features. In *ICCV*, pages 1150–1157, 1999. 4

[12] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16:187–198, 1997. 2

[13] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister. Statistics of infrared images. In *CVPR*, 2007. 2, 4

[14] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996. 4

[15] C. Rasmussen and C. Williams. Gaussian processes for machine learning. 2006. 3

[16] L. Wang and D.-C. He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910, 1990. 4

[17] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003. 2