

Fusing Images with Different Focuses Using Support Vector Machines

Shutao Li, James T. Kwok, Ivor W. Tsang, Yaonan Wang

Abstract—Many vision-related processing tasks, such as edge detection, image segmentation and stereo matching, can be performed more easily when all objects in the scene are in good focus. However, in practice, this may not be always feasible as optical lenses, especially those with long focal lengths, only have a limited depth of field. One common approach to recover an everywhere-in-focus image is to use wavelet-based image fusion. First, several source images with different focuses of the same scene are taken and processed with the discrete wavelet transform (DWT). Among these wavelet decompositions, the wavelet coefficient with the largest magnitude is selected at each pixel location. Finally, the fused image can be recovered by performing the inverse DWT. In this paper, we improve this fusion procedure by applying the discrete wavelet frame transform (DWFT) and the support vector machines (SVM). Unlike DWT, DWFT yields a translation-invariant signal representation. Using features extracted from the DWFT coefficients, a SVM is trained to select the source image that has the best focus at each pixel location, and the corresponding DWFT coefficients are then incorporated into the composite wavelet representation. Experimental results show that the proposed method outperforms the traditional approach both visually and quantitatively.

Index Terms—Image fusion, Wavelet transform, Support vector machines.

I. INTRODUCTION

OPTICAL lenses suffer from the problem of limited depth of field. According to the lens formula, only objects at one particular depth will be truly in focus. Consequently, if one object in the scene is in focus, another object at a different distance from the lens will be out of focus and thus blurred. The degree of this blurring is affected by a number of factors, including the object distance, the focal length and f -number of the lens, and the distance between the lens and the sensor plane [1]. On the other hand, many vision-related processing tasks, such as edge detection, image segmentation and stereo matching, can be more easily performed on focused images than on defocused ones. Hence, it is often advantageous if an everywhere-in-focus image can be recovered [2].

One approach to address this problem is by first estimating from the image the distance between the sensor and the objects in the scene. A number of depth computation techniques based on focus blur information, such as depth from focus [3], [4] and depth from defocus [5], [1], [6] methods, have been proposed. An everywhere-in-focus image can then be

recovered by deconvoluting one of the blurred images [2]. However, this approach typically demands a knowledge of the camera parameters and/or a model of the camera point spread function, which thus leads to a lot of mundane camera calibration.

In this paper, we follow another approach by performing image fusion [7], [8], [9], [10]. For the same object scene, two or more pictures (source images) are taken from the same camera position but with different focuses, such that each relevant object is in focus in at least one of them. These images are then fused, with the hope that all the objects will become in focus in the resultant image. In general, depending on the stage at which the combination mechanism takes place, image fusion can be divided into three categories, namely, pixel level, feature level and decision level [11]. Pixel level fusion works directly on the pixels obtained at the sensors' outputs. Feature level fusion, on the other hand, works on image features extracted from the source images. Decision level fusion works at an even higher level, and merges the interpretations of different images obtained after image understanding. Both feature level and decision level fusion may involve loss of information in the information extraction process, which consequently leads to less accurate fusion results [11]. In this paper, we will focus on pixel level fusion.

The simplest pixel level image fusion method just takes the pixel-by-pixel average of the source images. This, however, often produces undesirable side effects such as reduced contrast. A more successful method that has been explored in recent years is by using multiscale transforms. Examples include the Laplacian pyramid [12], [13], gradient pyramid [7], ratio-of-low-pass pyramid [14], morphological pyramid [15] and the discrete wavelet transform (DWT) [16], [8], [17], [10]. The basic idea is to perform multiresolution decomposition on each source image, and then integrate all these decompositions to obtain one composite representation, from which the fused image can be recovered by performing the corresponding inverse transform. However, many of these multiresolution decompositions are not translation-invariant because of an underlying down-sampling process [18]. Hence, in practice, their performance quickly deteriorates when there is slight object movement or when the source images cannot be perfectly registered. One way to alleviate this problem is by using the discrete wavelet frame transform (DWFT) [19], [20]. DWFT closely resembles the DWT, but uses an overcomplete wavelet decomposition by avoiding DWT's underlying down-sampling process. Its resultant signal representation is thus both aliasing free and translation-invariant.

As coefficients with large magnitudes correspond to salient

Shutao Li and Yaonan Wang are with the College of Electrical and Information Engineering, Hunan University, Changsha, People's Republic of China. James T. Kwok and Ivor W. Tsang are with the Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Email: shutao.li@yahoo.com.cn, [jamesk,ivor]@cs.ust.hk, yaonan@mail.hunnu.edu.cn.

features in the image, the *choose-max* rule is usually employed to fuse the multiresolution decompositions together [8]. To be more specific, corresponding coefficients from all decompositions are compared and the one with the largest magnitude is selected for use in the composite representation. However, this simple rule obviously may not produce optimal results.

In this paper, we use DWFT for the multiresolution decomposition, and then replace the choose-max rule by support vector machines (SVM) [21] for fusing the wavelet coefficients. The SVM is a recent classification technique that has outperformed many conventional approaches in various applications (e.g., [22], [23]). Given a training set containing m tuples $\{(\mathbf{x}_j, y_j)\}_{j=1}^m$ (where \mathbf{x}_j is the input and y_j is the corresponding output), the SVM first maps the patterns from input space to some feature space, and then separates the different classes by constructing a maximum margin hyperplane. Mathematically, this involves solving the quadratic programming problem:

$$\text{maximize } \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j),$$

w.r.t. the α_i 's, subject to the constraints

$$\sum_{i=1}^m \alpha_i y_i = 0, \quad \text{and } C \geq \alpha_i \geq 0, \quad (i = 1, \dots, m). \quad (1)$$

Here, C is a user-defined regularization parameter. Thus, SVMs have the important computational advantage that no nonlinear optimization is involved. Moreover, unlike other machine learning methods, its performance is related not to the input dimensionality, but to the margin with which it separates the data. Interested readers may consult [24], [25], [21] for more details.

The rest of this paper is organized as follows. Section II describes our proposed fusion scheme. Experimental results are then presented in Section III, and the last section gives some concluding remarks.

II. WAVELET-BASED FUSION SCHEMES

In this Section, we first give a brief introduction on the general framework of wavelet-based image fusion and introduce some abbreviations (Table I) that will be used in Section III. Interested readers may consult the survey in [10] for more details. In the following, we assume that the source images have already been registered, using either global [26], [27] or local motion estimation techniques [12]¹. The fusion procedure then takes the following steps:

- 1) Obtain the wavelet transform of each source image. Here, DWT, DWFT or other wavelet transform appropriate to the application can be used.
- 2) Measure an activity level from the wavelet coefficients, by treating each coefficient separately (*coefficient-based activity* (CBA)) or by averaging over a small window at each coefficient location (*window-based activity*

¹Recently, Zhang and Blum [28] proposed a hybrid registration method that is particularly suitable for image fusion applications. In cases where the source images cannot be perfectly registered, preliminary experimental results show that the proposed method is still quite robust [29].

(WBA)). WBA can be further classified as WA-WBA (*weighted-average WBA*), in which a weighted average is used, or RF-WBA (*rank-filter WBA*), which simply picks the largest coefficient inside the window.

- 3) Combine the multiple sets of wavelet coefficients together. The most popular method is to select the coefficient with the largest activity level at each pixel location (*choose-max* (CM)). Alternatively, one can also take a weighed average of the different sets of coefficients (*weighted average* (WA)), where the weights are determined by the activity levels of the source coefficients.
- 4) Optionally, perform consistency verification, which ensures that a fused coefficient does not come from a different source image from most of its neighbors. Usually, this is implemented by using a small majority filter (*window-based verification* (WBV)). In the sequel, NV denotes that no consistency verification is performed.
- 5) Use the inverse wavelet transform to recover the fused image.

From a pattern recognition viewpoint, the fusion task in Step 3 can be considered as a classification problem, namely, the selection of the source image with the best focus at a particular pixel location. With the development of more advanced classifiers like the SVM, one expects much room for improvement over the simple CM and WA schemes mentioned above. In the following, we first consider fusing two monochrome source images A and B . The proposed procedure (Figure 1) follows the general framework as described above:

- 1) Decompose the two source images by DWFT to d levels, resulting in a total of $3d$ details subbands and one approximation subband. As the approximation subband is a low-pass filtered version of the original image, it will contain little edge information after a sufficient number of decompositions and thus cannot help in deciding the clarity of the source image. Hence, this approximation subband will not be used in constructing the feature vector for SVM (though it will still be used in reconstructing the fused image). In the sequel, we denote the wavelet coefficient for image A (or B) at position (i, j) of the $b_1 b_2$ details subband (where $b_1 b_2$ can be either HL, LH or HH) at decomposition level k by $w_{b_1 b_2, k}^{(A)}(i, j)$ (or $w_{b_1 b_2, k}^{(B)}(i, j)$).
- 2) For each details subband, compute an activity level at each pixel location by CBA, WA-WBA or RF-WBA. Notice that this can be defined for *every* pixel location at all levels because all details subbands of DWFT are of the same size as the input image. Denote the resulting activity levels for A and B by $A_{b_1 b_2, k}(i, j)$ and $B_{b_1 b_2, k}(i, j)$ respectively.
- 3) Train a SVM to determine whether coefficients from A or B should be used at position (i, j) . The inputs to the SVM are the activity levels obtained from Step 2, while the target output label is 1 if A is considered clearer than B at (i, j) , and -1 otherwise (supervised learning). As there are only two source images here, we will simply use the difference vector $\{A_{HL, k}(i, j) - B_{HL, k}(i, j), A_{LH, k}(i, j) - B_{LH, k}(i, j), A_{HH, k}(i, j) -$

TABLE I
A SUMMARY OF THE ABBREVIATIONS USED.

Abbreviation	Description
CM	choose-max coefficient combining
CBA	coefficient-based activity level
DWFT	discrete wavelet frame transform
DWT	discrete wavelet transform
MLP	multilayer perceptron
NV	no consistency verification
RBF	radial basis function
SVM	support vector machine
WA	weighted average coefficient combining
WBA	window-based activity measurement
RF-WBA	WBA using rank filter
WA-WBA	WBA using weighted-average
WBV	window-based consistency verification

$B_{HH,k}(i, j)\}_k$ as input.

- 4) Perform testing over the whole image using the trained SVM. If the SVM output, $out(i, j)$, at position (i, j) is positive, coefficients for all the details and approximation subbands of the fused image at this pixel location will come from A , and vice versa. In other words, the fused coefficient at (i, j) is given by

$$w_{b_1 b_2, k}(i, j) = \begin{cases} w_{b_1 b_2, k}^{(A)}(i, j) & \text{if } out(i, j) > 0, \\ w_{b_1 b_2, k}^{(B)}(i, j) & \text{otherwise,} \end{cases}$$

for all subbands $b_1 b_2 \in \{LL, HL, LH, HH\}$ and all decomposition levels k .

- 5) Optionally, perform consistency verification on the result obtained in Step 4.
- 6) Finally, the fused image is recovered by performing the inverse DWFT on the fused coefficients.

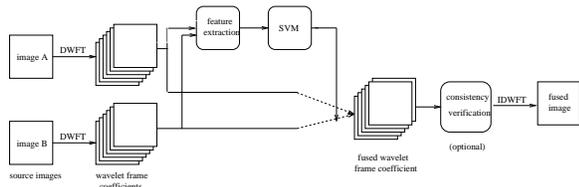


Fig. 1. Schematic diagram of the proposed fusion method.

In situations with more than two source images, the binary classification problem in Step 3 becomes a multi-class classification problem. In the SVM literature, a common approach is to cast this as a number of binary classification problems [21]. In other words, we construct one SVM for each source image, which decides whether this source image has the best focus at the pixel under consideration. In Step 4 above, the binary classifier with the highest output will be selected as winner, and all subband coefficients from the corresponding source image will be assigned to this pixel.

For color images, they can first be represented in the RGB, YUV or other color spaces, and DWFT is then performed on each color component. However, the components in some

color spaces (such as the RGB) are highly correlated. Hence, if these components are fused independently, undesirable visual effects may occur [12]. To alleviate this problem, we can first concatenate the feature vectors (obtained in Step 3) for all color components, and then feed this composite vector into the SVM for training and testing.

III. EXPERIMENTS

A. Setup

In this Section, we perform wavelet-based image fusion on five sets of 256-level images: *Balloon* (of size 480×640), *Clock* (512×512), *Lab* (480×640), *Pepsi* (512×512) and *Disk* (480×640). Because of the lack of space, only images for *Balloon* and *Clock* (Figures 2) will be shown in the sequel. For *Balloon*, the reference (everywhere-in-focus) image in Figure 2(a) is available. We then obtain Figures 2(b) and 2(c) by blurring its left and right halves respectively. The center of blurring is chosen to be at the clearest region of the reference image, so as to simulate the effect of finite depth of field in typical cameras. On the other hand, the other images (*Clock*, *Lab*, *Pepsi* and *Disk*) contain multiple objects at different distances from the camera, and are thus naturally blurred when taken. In these cases, the reference images are obtained manually by using Adobe Photoshop. Experiments have also been performed on color images with a similar performance gain as for monochrome images, and so will not be reported here.

A number of wavelet-based fusion schemes, with different combinations of wavelet transforms (DWT and DWFT), activity level measurements (CBA, WA-WBA and RF-WBA), coefficient combining methods (CM, WA, SVM and multilayer perceptron (MLP) [30]) and consistency verification methods (NV and WBV) are compared. For both wavelet transforms, we use the biorthogonal B-spline wavelet basis (with the analysis and synthesis filter coefficients shown in Table II) and three levels of decomposition. For WA-WBA and RF-WBA, a 3×3 window with the weights $\begin{bmatrix} \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{2}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix}$ is

used. For the SVM, we use the Matlab SVM toolbox [31]² with the linear kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ (where $\mathbf{x}_i, \mathbf{x}_j$ are input vectors) and the radial basis function (RBF) kernel (also called the Gaussian kernel) $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where we set $\frac{1}{\beta} = \frac{1}{N(N-1)} \sum_{i,j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|^2$ (N being the number of training patterns). As for the MLP, we use the two-layer feedforward network implementation from the Netlab neural network software³. Preliminary experiments suggest using 40 tanh hidden units and a logistic output unit. Training is performed by using the scaled conjugate gradient algorithm for 1000 iterations, and the fusion procedure is the same as that for SVM. All the schemes are implemented in Matlab and run on an AMD-K7 1GHz machine with 1G RAM.

Unlike the coefficient combining methods of CM and WA, both SVM and MLP require a training phase. In this paper, we train only one SVM (and one MLP) using patterns extracted from the *Balloon* image, and then use the trained classifier to perform fusion on all image sets. In practical implementations, this means that the trained classifier can simply be distributed as part of the software and no mundane training by the user is required. The training patterns are extracted from two pairs of 30×30 windows from the two *Balloon* source images. The first source image is clearer in one pair of the windows, while the second source image is clearer in the other pair. Each pixel generates one training pattern, and so the training set has a total of 1800 training patterns.

Three criteria are used to compare the fusion results. The first criterion is the root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (R(i, j) - F(i, j))^2},$$

where F is the fused image and R is the reference image (both are of size $M \times N$). The second criterion is the mutual information

$$\text{MI} = \sum_{i_1=1}^L \sum_{i_2=1}^L h_{R,F}(i_1, i_2) \log_2 \frac{h_{R,F}(i_1, i_2)}{h_R(i_1)h_F(i_2)},$$

where $h_{R,F}$ is the normalized joint gray level histogram of images R and F , h_R, h_F are the normalized marginal histograms of the two images, and L is the number of gray levels. Notice that MI measures the reduction in uncertainty about the reference image due to the knowledge of the fused image, and so a larger MI is preferred. The last criterion is the CPU time (in seconds) required by Matlab for the whole fusion procedure.

The SVM requires the setting of a regularization parameter (C in (1)) that trades off the margin with training errors. Figure 3 shows some preliminary results on the effects of C on RMSE and MI (when the SVM is used with the NV+CBA setting on the *Balloon* image). As can be seen, the performance is relatively stable over a large range of C . Thus, for simplicity, we will always use the value of $C = 1000$ in the experiments.

²The Matlab SVM toolbox can be downloaded from <http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox/>.

³The Netlab neural network software can be downloaded from <http://www.ncrg.aston.ac.uk/netlab/>.



(a) *Balloon*.



(b) focus on the right.



(c) focus on the left.



(d) *Clock*.



(e) focus on the left clock.



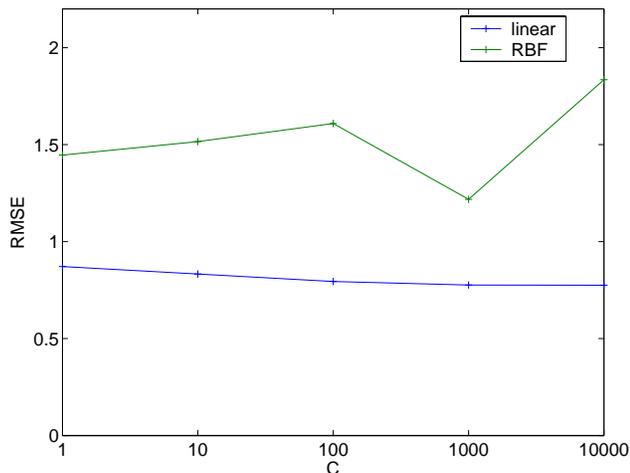
(f) focus on the right clock.

Fig. 2. The everywhere-in-focus reference images and blurred source images of *Balloon* and *Clock*. The training set for SVM/MLP is selected from the two windows marked by the rectangles in Figures 2(b) and 2(c).

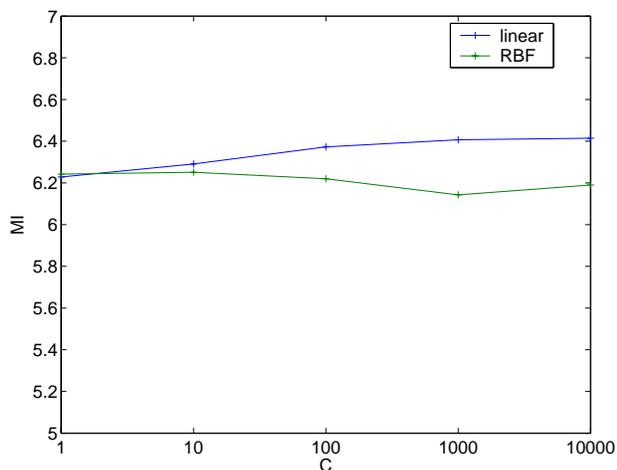
TABLE II

COEFFICIENTS FOR THE ANALYSIS (h) AND SYNTHESIS (\tilde{h}) FILTERS OF THE BIORTHOGONAL B-SPLINE WAVELET BASIS.

n	0	± 1	± 2	± 3	± 4
h	0.602949	0.266864	-0.078223	-0.016864	0.026749
\tilde{h}	0.557543	0.295636	-0.028772	-0.045636	0



(a) RMSE.



(b) MI.

Fig. 3. Variations of RMSE and MI at $C = 1, 10, 100, 1000, 10000$.

B. Fusion Results

Tables III and IV compare the image fusion schemes based on RMSE and MI. Because of the lack of space, detailed results are reported only for the SVM using the linear kernel. For the others (i.e., SVM with RBF kernel, MLP and the traditional DWT and DWFT schemes), we have only included the best result over all the tested fusion schemes. As can be seen, the linear SVM shows marked improvements over the others. Moreover, its performance is quite insensitive to the particular setting used for consistency verification and

activity level measurement. SVM with the RBF kernel also outperforms traditional DWT and DWFT schemes on the *Balloon* images. However, its performance is significantly inferior to that of the linear SVM, especially on application to the other image sets. This can be attributed to the use of DWFT coefficients in measuring activity levels, which, as demonstrated in previous studies, produces good input features for the SVM in deciding which source image has the best focus. Consequently, the extra nonlinearity associated with a nonlinear kernel does not help in most cases. On the other hand, MLP is also a powerful classifier and outperforms DWT and DWFT in general. However, unlike the SVM, it lacks the margin maximizing property and thus does not perform as well. Furthermore, our relative performance on the various fusion settings are mostly in line with those reported in [10]. For example, both WA-WBA and RF-WBA perform better than CBA and RF-WBA performs better than WA-WBA. On the other hand, while WBV performed slightly better than NV in [10], here they have almost identical performance under most combinations.

Because of the large number of combinations that have been performed, here we only show the fused images obtained by the linear SVM with NV+WA-WBA setting (Figure 4). The fused images produced by the linear SVM are basically a combination of the in-focus parts of the source images, while results based on the other methods are much inferior.

Finally, Table V compares the computational speeds of the various fusion schemes. As expected, the simplest DWT scheme, with no consistency verification and no averaging in measuring the activity level, is almost always the fastest. However, as we have seen earlier, its performance in terms of RMSE and MI are much inferior to those of the linear SVM. On the other hand, notice that the linear SVM, with the NV+WA-WBA setting, takes only about twice the time required by the fastest scheme. Recall that its performance is comparable with that of the best setting in Tables III and IV, our recommendation is thus to use the linear SVM with NV+WA-WBA in this fusion application.

IV. CONCLUSION

In this paper, we study the use of DWFT and SVM for fusing images with different focuses of the same scene in order to obtain an everywhere-in-focus image. DWFT is advantageous over DWT in that the decomposed signal representation is translation-invariant, while SVM outperforms standard coefficient combining methods in finding the source image with the best focus at a particular pixel location. Experiments demonstrate that the proposed fusion scheme, particular with the use of the linear SVM, outperforms a number of conventional schemes both visually and quantitatively. By working on

TABLE III

RMSE'S OBTAINED BY THE VARIOUS FUSION SCHEMES (THE LOWEST VALUES ARE INDICATED IN BOLD).

scheme			<i>Balloon</i>	<i>Clock</i>	<i>Lab</i>	<i>Pepsi</i>	<i>Disk</i>
linear SVM	NV	CBA	0.78	5.12	2.85	3.14	3.94
		WA-WBA	0.76	5.11	2.85	3.10	3.94
		RF-WBA	0.75	4.99	2.90	3.30	3.96
	WBV	CBA	0.74	5.13	2.85	3.09	3.98
		WA-WBA	0.74	5.17	2.86	3.11	3.95
		RF-WBA	0.74	4.99	2.90	3.32	3.95
RBF SVM			1.04	4.98	6.19	4.32	9.43
MLP			0.88	5.71	2.76	3.27	3.99
DWT			1.14	6.95	4.24	3.95	5.66
DWFT			1.34	6.92	4.85	4.00	6.46

TABLE IV

MI'S OBTAINED BY THE VARIOUS FUSION SCHEMES (THE HIGHEST VALUES ARE INDICATED IN BOLD).

scheme			<i>Balloon</i>	<i>Clock</i>	<i>Lab</i>	<i>Pepsi</i>	<i>Disk</i>
linear SVM	NV	CBA	6.41	4.69	4.76	4.44	4.62
		WA-WBA	6.45	4.76	4.83	4.61	4.70
		RF-WBA	6.47	4.94	4.86	4.71	4.77
	WBV	CBA	6.51	4.89	4.84	4.66	4.71
		WA-WBA	6.52	4.85	4.88	4.72	4.76
		RF-WBA	6.51	5.01	4.89	4.76	4.82
RBF SVM			6.48	5.01	4.53	4.49	4.00
MLP			6.27	4.72	4.85	4.68	4.74
DWT			6.18	4.16	4.09	3.89	3.76
DWFT			5.77	3.75	4.03	3.76	3.47

TABLE V

NUMBER OF CPU SECONDS REQUIRED BY MATLAB FOR THE VARIOUS FUSION SCHEMES (THE SHORTEST TIME IS INDICATED IN BOLD). RECALL THAT SVM TRAINING IS ONLY REQUIRED FOR THE *Balloon* IMAGE.

scheme			<i>Balloon</i>	<i>Clock</i>	<i>Lab</i>	<i>Pepsi</i>	<i>Disk</i>
linear SVM	NV	CBA	29.76	22.32	26.62	24.04	28.01
		WA-WBA	28.41	21.48	25.48	23.80	26.95
		RF-WBA	233.96	194.27	227.94	195.59	242.32
	WBV	CBA	43.03	33.08	39.56	35.72	48.03
		WA-WBA	41.31	32.80	38.66	35.38	45.07
		RF-WBA	244.88	208.36	242.66	206.85	311.93
RBF SVM			30.73	23.17	27.46	25.34	35.24
MLP			209.83	26.89	31.30	30.54	31.70
DWT			11.78	15.09	22.43	12.32	13.51
DWFT			59.89	61.83	65.56	58.48	63.73

the fused image rather than on the original defocused image, vision-related processing tasks such as edge detection, image segmentation and stereo matching can be expected to yield more accurate results. Moreover, in practical implementations, the trained SVM can be conveniently distributed as part of the software and does not require mundane training by the user.

In the future, we plan to extend the proposed scheme to other fusion tasks. One direction is for the fusion of images with different dynamic ranges [12]. Here, multiple images of the same scene are taken, and each may contain regions that

are either over-saturated or under-saturated. The classification problem for SVM then becomes selecting the source image with the best illumination at a particular pixel location. Other possibilities include the fusion of heterogeneous images, such as fusing the high spatial resolution SPOT panchromatic images with the lower resolution, multispectral Landsat Thematic mapping (TM) images in remote sensing applications [32].



(a) Balloon.



(b) Clock.

Fig. 4. Fused images obtained by the linear SVM, with NV+WA-WBA.

ACKNOWLEDGMENTS

This research has been partially supported by the Research Grants Council of the Hong Kong Special Administrative Region under grants HKUST2033/00E and HKUST6195/02E. The authors would like to thank Rick Blum for providing some of the images used in the experiments.

REFERENCES

- [1] D.T. Pham and V. Aslantas, "Depth from defocusing using a neural network," *Pattern Recognition*, vol. 32, pp. 715–727, 1999.
- [2] M. Subbarao, T.-C. Wei, and G. Surya, "Focused image recovery from two defocused images recorded with different camera settings," *IEEE Transactions on Image Processing*, vol. 4, no. 12, pp. 1613–1628, 1995.
- [3] J. Ens and P. Lawrence, "Investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 97–108, 1993.
- [4] M. Subbarao and T. Choi, "Accurate recovery of three dimensional shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 266–274, 1995.
- [5] S.H. Lai, C.W. Fu, and S. Chang, "A generalized depth estimation algorithm with a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 405–411, 1992.
- [6] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *International Journal Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.
- [7] P.T. Burt and R.J. Kolczynski, "Enhanced image capture through fusion," in *Proceedings International Conference on Computer Vision*, 1993, pp. 173–182.
- [8] H. Li, B.S. Manjunath, and S.K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, May 1995.

- [9] W.B. Seales and S. Dutta, "Everywhere-in-focus image fusion using controllable cameras," *Proceedings of SPIE*, vol. 2905, pp. 227–234, 1996.
- [10] Z. Zhang and R.S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1315–1326, Aug. 1999.
- [11] P.K. Varshney, "Multisensor data fusion," *Electronics and Communication Engineering Journal*, pp. 245–253, Dec. 1997.
- [12] L. Bogoni and M. Hansen, "Pattern-selective color image fusion," *Pattern Recognition*, vol. 34, pp. 1515–1526, 2001.
- [13] P.T. Burt and E.H. Andelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, pp. 532–540, 1983.
- [14] A. Toet, L.J. van Ruyven, and J.M. Valetton, "Merging thermal and visual images by a contrast pyramid," *Optical Engineering*, vol. 28, no. 7, pp. 789–792, July 1989.
- [15] G.K. Matsopoulos, S. Marshall, and J.N.H. Brunt, "Multiresolution morphological fusion of MR and CT images of the human brain," *Proceedings of IEE: Vision, Image and Signal Processing*, vol. 141, no. 3, pp. 137–142, 1994.
- [16] L.J. Chipman, Y.M. Orr, and L.N. Graham, "Wavelets and image fusion," in *Proceedings of the International Conference on Image Processing*, Washington, USA, 1995, pp. 248–251.
- [17] D.A. Yocky, "Image merging and data fusion by means of the discrete two-dimensional wavelet transform," *Journal of the Optical Society of America A: Optics, Image Science and Vision*, vol. 12, no. 9, pp. 1834–1841, 1995.
- [18] S.G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 2nd edition, 1999.
- [19] A. Laine and J. Fan, "Frame representations for texture segmentation," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 771–780, 1996.
- [20] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.
- [21] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [22] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proceedings of Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.
- [23] B. Schölkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, Nov. 1996.
- [24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.
- [25] B. Schölkopf and A.J. Smola, *Learning with Kernels*, MIT University Press, Cambridge, MA, 2002.
- [26] R.H.J.R. Bergen, P.J. Burt, and S. Peleg, "Computing two motions from three frames," in *Proceedings of the IEEE International Conference on Computer Vision*, 1990, pp. 27–32.
- [27] G.W.M. Hansen, K. Dana, and P. Burt, "Real-time scene stabilization and mosaic construction," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 1994, pp. 54–62.
- [28] Z. Zhang and R.S. Blum, "A hybrid image registration technique for a digital camera image fusion application," *Information Fusion*, vol. 2, no. 2, pp. 135–149, June 2001.
- [29] S. Li, J.T. Kwok, and Y. Wang, "Fusing images with multiple focuses using support vector machines," in *Proceedings of the International Conference on Artificial Neural Networks*, Madrid, Spain, Aug. 2002, pp. 405–410.
- [30] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford, New York, 1995.
- [31] G.C. Cawley, "MATLAB support vector machine toolbox," 2000, University of East Anglia, School of Information Systems, Norwich, Norfolk, U.K. NR4 7TJ.
- [32] S. Li, J.T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images," *Information Fusion*, vol. 3, no. 1, pp. 17–23, 2002.

PLACE
PHOTO
HERE

Shutao Li was born in 1972 in P.R. China. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Hunan University, Changsha, P.R. China in 1995, 1997, and 2001, respectively. He joined the College of Electrical and Information Engineering, Hunan University, in 2001. He was with Department of Computer Science, Hong Kong University of Science and Technology, from May 2002 to October 2002. He is currently a Research Staff Member with the Department of Computer Science, Royal Holloway College, University of

London. His main research interests include image processing and machine learning.

PLACE
PHOTO
HERE

James Tin-Yau Kwok received the Ph.D. degree in computer science from the Hong Kong University of Science and Technology in 1996. He then joined the Department of Computer Science, Hong Kong Baptist University as an Assistant Professor. He returned to the Hong Kong University of Science and Technology in 2000 and is now an Assistant Professor in the Department of Computer Science. His research interests include kernel methods, machine learning, pattern recognition and artificial neural networks.

PLACE
PHOTO
HERE

Ivor Wai-Hung Tsang received his B.Eng. degree and M.Phil. degree in the Computer Science in 2001 and 2003 respectively from the Hong Kong University of Science and Technology (HKUST). He is currently pursuing the PhD degree in the HKUST. He was the Honor Outstanding Student in 2001. His scientific interests includes machine learning and kernel methods.

PLACE
PHOTO
HERE

Yaonan Wang received the B.S. degree in Computer Engineering from East China Technology Institute, China in 1981. He received the M.S. and Ph.D. degrees in Electrical Engineering from Hunan University, China, in 1990 and 1994, respectively. From 1994 to 1995 he was a Postdoctoral Research Fellow in National University of Defense Technology. From 1981 to 1994 he worked in East China Technology Institute. Since 1995, he has been a professor in Hunan University, China. His research interests are image processing, pattern recognition, and industrial

process control.