DISTRIBUTED IMAGE COMPRESSION FOR SENSOR NETWORKS USING CORRESPONDENCE ANALYSIS AND SUPER-RESOLUTION

Raymond Wagner, Robert Nowak, Richard Baraniuk

Department of Electrical and Computer Engineering, Rice University

ABSTRACT

We outline a distributed coding technique for images captured from sensors with overlapping fields of view in a sensor network. First, images from correlated views are roughly registered (relative to a sensor of primary interest) via a low-bandwidth data-sharing method involving image feature points and feature point correspondence. An area of overlap is then identified, and each sensor transmits a low-resolution version of the common image block to the receiver, amortizing the coding cost for that block among the set of sensors. Super-resolution techniques are finally employed at the receiver to reconstruct a high-resolution version of the common block.

We discuss the registration and super-resolution techniques used and present examples of each step in the proposed coding process. A numerical analysis illustrating the potential coding benefit follows, and we conclude with a brief discussion of the key issues remaining to be resolved on the path to coder robustness.

1. OVERVIEW

Development on the hardware side of sensor networks is advancing at such at rapid rate that high-dimensional data gathering and processing at local sensor nodes is now not only feasible but on the immediate horizon. Specifically, dense networks of wirelessly linked and camera-equipped sensors with onboard power and computing resources will soon be deployable. These sensors will need to share information about recorded images with each other and the networks' end users, activities which necessitate a significant amount of wireless communication traffic. Since communication power consumption typically dominates over computation power consumption by orders of magnitude, it follows that the techniques for managing information collected at the sensors must become more sophisticated. Repeated transmission of whole images will make short work of a sensor's power supply, effectively rendering it dead to the network far earlier then if it were simply gathering and reporting lower dimensional data. Thus, a technique for sharing image data in an efficient manner both between sensors and with the user must be developed. Fortunately, in a sufficiently dense sensor network, the cameras will have correlated fields of view. We discuss here a proposed method to enable a distributed encoding of these images in which mutual information among adjacent cameras is exploited to lighten the communication burden for sensors wishing to transmit their images.

This work is similar in spirit to the ongoing DISCUS (distributed source coding using syndromes) effort outlined in [1]. DISCUS aims to develop a distributed coding framework to realize the theoretical coding gains promised by the Slepian-Wolf coding theorem from information theory. Ideally, little or no information need be exchanged among correlated sensors during the encoding process while reaping the full benefit of the correlation at the decoder. Where DISCUS approaches this problem from a more information theoretic angle, we wish to explore an application-specific solution which does not lend itself so easily to a probabilistic analysis. As the uncertainty between neighboring images is not described with a conveniently analyzable joint process such as a Gaussian, straightforward application of the Slepian-Wolf theorem would prove difficult. Thus, we instead chose to approach the problem by fusing a variety of signal processing solutions to realize similar coding gains.

We are interested in investigating the utility of a new method for distributed image compression based on feature points, correspondence analysis, and image super-resolution techniques. The configuration of feature points, obtained for instance from the downsampled output of an edge detector, describes the configuration of the key edges in an image. Since these feature points are small in number, they can be efficiently shared amongst a number of sensors. Using a correspondence algorithm, a sensor of interest can efficiently determine not only the spatial correlations between its locally acquired image and those of neighbors, but also a transformation to warp neighboring images to its own. This transform provides a rough registration between images, allowing sensors to identify a common image block appearing in all images of the set. Each sensor can then downsample its version of the common block and send the resulting low resolution image to the receiver, which super-resolves the set of low-resolution images into a reconstruction of the original high-resolution block. Complexity is moved out of the network and into the receiver, with the result that in-network processing is simple yet powerful.

2. COLLABORATIVE IN-NETWORK COMPRESSION SCENARIO

Image a scenario such as that depicted in Figure 1, in which a number (N) of sensors share different looks at a common scene, with one particular camera having the perspective of most interest (call it the primary sensor). Assuming sufficiently dense placement of the sensors, the $N-1$ non-primary cameras each have some overlap between their fields of view and that of the primary camera. The intersection of these $N-1$ overlapping regions describes a common region of overlap, where each sensor sees the same image block in part of its field of view.

We can compute pair-wise registrations between the image from the primary camera and those from each of the $N-1$ non-primary cameras. These registrations allow us to transform the non-primary images into the frame of reference of the primary

Supported by DARPA, NSF, ONR, AFOSR, and an NSF graduate fellowship. Email: [rwagner,nowak,richb]@rice.edu Web: dsp.rice.edu

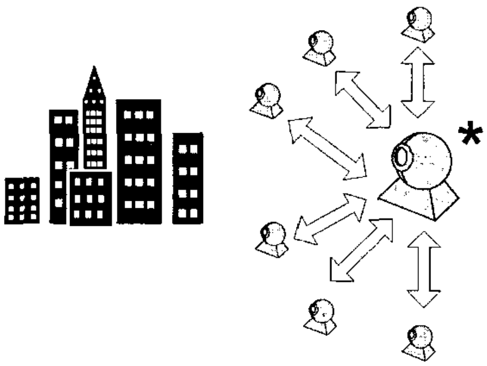


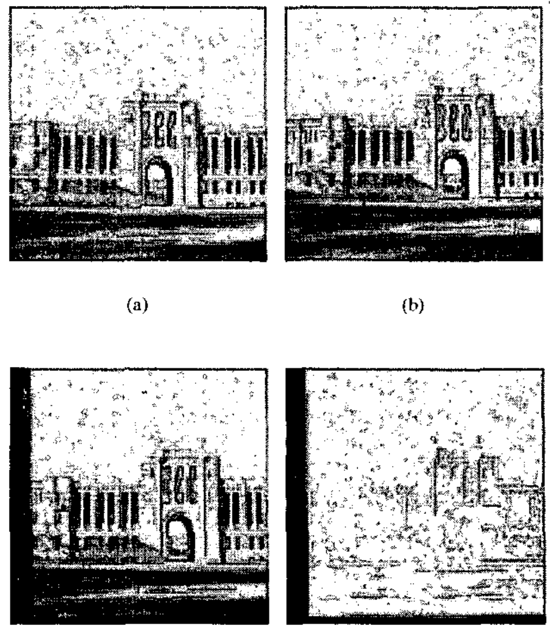
Fig. 1. Sensor network configuration, with primary sensor of interest (*) and other sensors with overlapping fields of view.

camera and identify the largest possible region of overlap as described above. Once this is accomplished, we can then amortize the coding cost for the common block among the entire set of sensors, requiring each to transmit a low-bitrate version of the block. This relieves a portion of the transmission duties of the primary sensor (the one fielding the most image download queries) and spreads the associated power loss among all the sensors in its neighborhood, thus prolonging its life and postponing the time when it drops out of the network due to depletion of its power supply. Such a loss would, of course, force the operator to resort to a non-preferred view of the scene, one in which potentially critical information is now out of the camera frame.

The approach proposed here is similar in spirit to that commonly used for stereo image compression, where similarities between two slightly displaced images are exploited to reduce the amount of information required for coding the pair [2], [3]. Whereas stereo compression registers a pair of images and then jointly encodes them, we wish to register multiple pairs of images (pairing each of the $N-1$ non-primary images with the primary) and then find an intelligent way to jointly encode the registered set. Of course, the stereo camera placement assumptions (inducing only small translational displacements) and image-sharing requirements (cameras typically have knowledge of both images of the pair) are also not valid in the context of our proposed scenario, requiring us to appeal to a more complex means of registration, as discussed in the next section.

3. REGISTRATION VIA CORRESPONDENCE ANALYSIS

To realize a low-bandwidth means of communicating image information among sensors, we employ the shape context algorithm proposed in [4]. This technique allows for the correspondence of similar images in terms only of their critical feature points via a shape descriptor known as the shape context. The process for corresponding a pair of images proceeds as follows and is repeated for registering each non-preferred sensor in the correlated set with the preferred sensor. First, both images are sampled to extract two sets of points which capture the critical features of the images. For example, these features can be pixels along dominant edges of the image. Any standard edge detection algorithm can be employed



(a) (b)

(c) (d)

Fig. 2. Example of feature point registration (a) original non-preferred image I_n (b) original preferred image I_p (c) warped I_n (d) difference image between I_p and warped sensor I_n

in this step. Then, in both sets, for each point a shape context is computed. The shape context is a log-polar histogram which describes the spatial relationship of all other points in the set to the sample point for which the context is computed. This yields two sets of sample points with associated shape contexts. A set of costs of matching each point in the first set to each in the second is computed via the χ^2 test statistic. Bi-partite graph matching is then employed to find the best one-to-one matching, with the option of matching only a fraction of the points by defining dummy points to which the rest (designated as outliers) can match. Given this set of correspondences, a transform which will warp one image into the frame of reference of the other can be estimated.

An example of the registration process between a pair of sensors is given in Figure 2. Two views of a scene (say, one of the $N-1$ non-preferred images I_n in Figure 2(a) and the preferred image I_p in Figure 2(b)) are captured by a neighboring pair of sensors. Following the shape context registration process, there exists a warping transform which can take I_n to the frame of reference of I_p, with the result being the image in Figure 2(c). To evaluate the effectiveness of this transformation, the difference image between the warped I_n and the original I_p is shown in Figure 2(d). Note that there exists a small translational/rotational error in the registration. Though this seems to be a problem, it will in fact allow us to realize the super-resolution endocing/decoding scheme described in the next section.

Two key benefits of such a registration scheme emerge. First, an extremely low-bandwidth dataset, the set of feature point locations in each image, is passed between pairs of sensors for the purpose of image correspondence. Second, this registration need be performed only once. Provided sensors and key static image features remain stationary following initial registration, sensor views should be related by the same set of warping transforms regardless of the amount of time passed since the first registration. Thus, the amount of bandwidth consumed in registration is extremely small compared to that used for the subsequent compression, especially as the number of image download requests increases. With this registration accomplished for each of the $N-1$ non-preferred images, we can now move on to exploiting the common area of image overlap which results.

4. DECODING VIA SUPER RESOLUTION

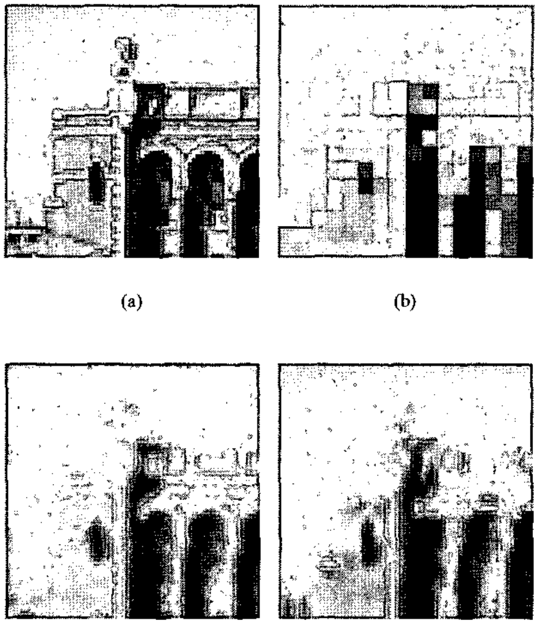
Image super-resolution is a rapidly developing field which tackles the problem of extracting a single high-resolution image from multiple low-resolution versions with unknown sub-pixel displacements induced by such processes as camera translation and rotation. An example of a classical problem in super-resolution is the extraction of a single high-resolution still image from multiple frames of a lower resolution video feed. For a good technical overview of problems and solutions in the domain of super-resolution, the reader is referred to [5].

The initial assumptions of the super-resolution problem fit quite naturally into our proposed distributed coding algorithm. In fact, whereas registration error would normally be a problem when exploiting image correlation, it now emerges as a key feature to enable super-resolving sensors' images at the receiver. Recall that the feature point correspondence step allows us to apply warping transforms to the set of available non-preferred images to register them with the preferred image. The images in this transformed set can then be intersected with each other to identify the largest common pixel block in the set (i.e. the largest block in the area of overlap). Once this common block has been identified in each transformed image, we have exactly the data required for super-resolution: a set of similar images related (through the registration error) by a set of translations and rotations.

We distribute the coding of this block among the group of sensors by requiring each to transmit a downsampled version of the post-registration block to the receiver. Thus, the coding cost for this block in the sensor of interest is spread among the entire neighborhood of sensors, reducing the transmission load on the preferred sensor and thus increasing its operational lifetime.

5. ANALYSIS OF POTENTIAL CODING GAIN

To illustrate the potential coding gain of such an approach, we implemented a basic version of the algorithm found in [6]. This choice of algorithm is by no means exclusive, as there exist a large number of solutions to the image super-resolution problem. For an exploration of the variety of common techniques, the interested reader is referred to the publication containing [5] The particular algorithm employed here applies Bayesian analysis techniques to extract unknown registration parameters from low-resolution images as follows. The unknown high-resolution image is modelled with a Gaussian prior, and low-resolution images are assumed to be generated from the original by multiplication with a transformation matrix W, which accounts for shifting, rotating, and downsampling the original via convolution with a Gaussian pointspread function centered on the low-resolution pixels. An expression for the probability of the set of low-resolution images given the registration parameters and unknown, original image is formulated. The high-resolution image is then marginalized out of this expression to yield one which only conditions the low-resolution images on the registration parameters. This marginal likelihood can then be iteratively maximized to yield estimates of the parameters. With these parameter estimates, the unknown high-resolution pixel values can be estimated by maximizing the conditional likelihood of the high-resolution image given the registration parameters and the low-resolution image set.



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

Fig. 3. Sample super-resolved images (a) original high-resolution image (b) sample low-resolution image (c) super-resolved image (d) zerotree-coded image using equivalent bitrate

For the purpose of our analysis, we synthetically generated a set of low-resolution test images from a 64x64 square-pixel detail (Figure 3(a)) of one of the images in Figure 2. Limiting our investigation to translational variations only (no rotational error) for ease of analysis, we generated a set of low-resolution images by applying to the test image different realizations of the aforementioned matrix W, calculated using random high-resolution pixel shifts, a downsampling factor of 4 in each direction, and constant pointspread function width. The resulting lo-resolution images were then wavelet zerotree coded using the method described in [7]. Figure 3(b) shows an example low-resolution image coded at a target bitrate of 4 bits per low-resolution pixel.

The super-resolved image was then reconstructed from this set of low-resolution images. First, the shift registration parameters were estimated as discussed above. Pointspread function width

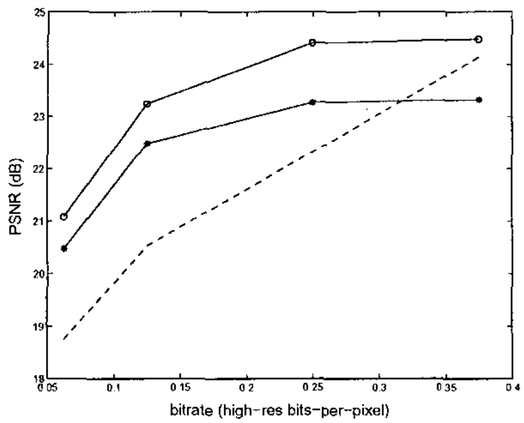


Fig. 4. Rate-distortion comparison using 16 sensors (o), 8 sensors (*), and zerotree encoding (dashed line)

was assumed known, as it would explicitly enter into the downsampling process conducted by the sensor before transmission in order to match the generative assumptions of the algorithm. The parameter estimates were then used to compute the maximum likelihood estimate of the high-resolution pixels (Figure 3(c)). Distortion of the reconstruction was then computed relative to the original using the peak signal-to-noise ratio (PSNR) [1] metric.

To evaluate the effectiveness of our method, distortion for the super-resolution reconstruction was compared against that from a simple zerotree encoding of the original high-resolution image. Zerotree bitrate values were chosen to represent a per-sensor bit budget equivalent to that used in each of the super-resolution experiments (Figure 3(d)). For instance, where a target bitrate of N bits-per-pixel was employed to encode the 16x16 low-resolution image for a total expenditure of 16^2N bits at that sensor, the comparison high-resolution image was encoded at a target bitrate of $16^2N/64^2$ bits-per-pixel (bpp). (The examples in Figure 3 used 4 bpp for the low resolution images and 0.25 bpp for the high-resolution zerotree encoding.) The distortion in PSNR for the various zerotree encodings was computed and compared against two families of super-resolution reconstructions.

Simulations were conducted using sets of 8 and 16 low-resolution images zerotree coded with target bitrates between 1 and 6 bits per low-resolution pixel. Results are shown in Figure 4, with PSNR in dB on the vertical axis and bitrate in high-resolution bpp on the horizontal axis. The curve for super-resolved image distortion using 16 low-resolution images is marked with circles, and that using 8 low-resolution images is marked with asterisks. When comparing points encoded at the same per-sensor bitrate, the super-resolution curves clearly demonstrate lower distortion (higher PSNR) than the plain zerotree coding over the majority of the interval examined, with the increase in performance becoming more pronounced as the bitrate of the high-resolution zerotree coding is reduced. Alternately, fixing distortion and examining bitrate shows us that the super-resolved images can yield comparable quality for a smaller per-sensor bit budget. The reconstruction using 8 images naturally does not perform as well as that using 16, as there is less information to exploit in the super-resolving process. Regardless, this exercise demonstrates that there is a definite coding gain to be realized using a distributed super-resolution coding method such as we have described.

[1] PSNR $= 10\log_{10}(M255^2/\text{MSE})$, where MSE is the mean squared error between the estimate and original and M is the number of data points in each image.

6. CONCLUSIONS

We have outlined a method for distributed coding of images in a sensor network by exploiting correlation in neighboring sensors' images. Beginning with a crude registration based on low-bandwidth feature point exchange and correspondence, super-resolution techniques can allow the each sensor in the correlated set to transmit a low-resolution version of the overlap area by exploiting small errors in the registration process. This effectively amortizes the coding cost for that common block among all the sensors, relieving any particular sensor of interest from the task of transmitting the whole block. We have demonstrated that there exists a coding gain to be exploited by such a scheme, encouraging further development of the method.

A number of issues still remain to be resolved with this technique. Our model only considers registrations involving simple translations and rotations. This effectively restricts the placement of sensors in the neighborhood to the same family of displacements. A more realistic model would incorporate affine transforms. It remains to be seen how affine registration error will affect the super-resolution model under consideration. Additionally, our experiences with the super-resolution algorithm seem to indicate that a relatively large number of low-resolution images are needed to realize a reconstruction of acceptable quality. This imposes additional restrictions on the distribution of sensors with which such a coding scheme can be realized.

7. REFERENCES

[1] S.S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, Mar. 2002.

[2] N.V. Boulgouris and M.G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 898–903, Oct. 2002.

[3] J. Xu, Z. Xiong, and S. Li, "High performance wavelet-based stereo image coding," in *IEEE International Symposium on Circuits and Systems*, 2002, pp. 273–276.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, Apr. 2002.

[5] M.G. Kang S.C. Park, M.K. Park, "Super-resolugion image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, May 2003.

[6] M.E. Tipping and C.M. Bishop, "Bayesian image super-resolution," *Advances in Neural Information Processing Systems*, 2002, pending publication.

[7] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.