Single Image Example-Based Super-Resolution Using Cross-Scale Patch Matching and Markov Random Field Modelling

Tijana Ružić, Hiêp Q. Luong, Aleksandra Pižurica, and Wilfried Philips

Ghent University, TELIN-IPI-IBBT,
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
tijana.ruzic@telin.ugent.be

Abstract. Example-based super-resolution has become increasingly popular over the last few years for its ability to overcome the limitations of classical multi-frame approach. In this paper we present a new example-based method that uses the input low-resolution image itself as a search space for high-resolution patches by exploiting self-similarity across different resolution scales. Found examples are combined in a high-resolution image by the means of Markov Random Field modelling that forces their global agreement. Additionally, we apply back-projection and steering kernel regression as post-processing techniques. In this way, we are able to produce sharp and artefact-free results that are comparable or better than standard interpolation and state-of-the-art super-resolution techniques.

Keywords: Super-resolution, self-similarities, Markov Random Field, kernel regression

1 Introduction

Super-resolution (SR) plays an important role in image processing applications nowadays due to the huge amount of low resolution video and image material. Low resolution is a consequence of using low-cost imaging sensors for image/video acquisition, such as webcams, cell phones and surveillance cameras. Furthermore, the increasing popularity of HDTV makes the SR methods necessary for resolution enhancement of NTSC and PAL recordings.

The task of SR is to infer a high-resolution image from one or more low resolution images. Among many SR techniques, two approaches can be identified: classical and example-based approach. Classical SR methods attempt to reconstruct a high-resolution (HR) image from a sequence of degraded low-resolution (LR) images taken from the same scene at sub-pixel shifts [1, 2]. Each output pixel is related to one or more input pixels by the acquisition or degradation model. If there is an insufficient number of LR images, prior knowledge can be used as an additional source of information. Classical SR in practice, however, is limited to the the magnification factor smaller than two [4]. Example-based

SR is able to overcome this limitation. The goal of the example-based approach [3, 4] is to fill in the missing high frequencies by searching for highly similar patches in the external database that also contains high-resolution information. The method actually consists of two steps: a learning and a reconstruction step. The former involves searching for k nearest neighbours in the database for each LR patch of the input image, while the latter combines the corresponding HR patches of those nearest neighbours to form the HR image.

A problem with example-based methods is that they involve storing and searching large databases. Searching the database can be avoided by using it only to learn the interpolation functions [5, 6], but still this external database is necessary. Additionally, it is not guaranteed that the database contains the true high-resolution details which may cause the so called "hallucination" effect. Furthermore, this database needs to be large enough to provide good results which makes learning or searching computationally more demanding.

A solution to the previously mentioned problems is to use the LR input image itself as a search space in the learning phase, as implied originally in [3]. Based on this idea, several single image super-resolution techniques have been developed [7, 8]. In [8], all examples are obtained by searching for nearest neighbours within the Gaussian pyramid of the input LR image. This example-based part is combined with the classical SR approach to yield an HR image with an arbitrary magnification factor. The use of a single image is justified by the level of patch redundancy within the same scale and across different levels of Gaussian pyramid. Following this reasoning, we have also developed a single-image super-resolution algorithm that, in addition to these non-local similarities within and across scales, uses sparsity constraints to perform image super-resolution [9].

In this paper we propose a novel single image example-based super-resolution algorithm which combines the learning phase of [8] by searching for examples within the Gaussian pyramid of the input image itself and the reconstruction phase of [3], which uses the Markov Random Field (MRF) model to reconstruct the HR image. The main benefit of such learning approach is that no external database is required which results in faster search and absence of "hallucination" effect (when compared with [3]). On the other hand, using MRF in the reconstruction enables us to stay in the example-based domain without combining it with classical SR as in [8]. There are a few advantages to this in comparison with [8]. First of all, we can use only one level of the pyramid as the search space whose sub-sampling factor corresponds to the magnification factor instead of multiple levels with non-integer sub-sampling factor and, thus, again decrease the computation time. Second, we reconstruct only the HR image of the desired resolution rather than employing course-to-fine reconstruction of images at intermediate resolutions. Finally, we avoid sub-pixel registration which often causes inaccurate results.

Another contribution of this paper is that we show that a simpler and faster method can be used for inference in MRF instead of belief propagation used originally in [3]. We use our method from [10] called neighbourhood-consensus
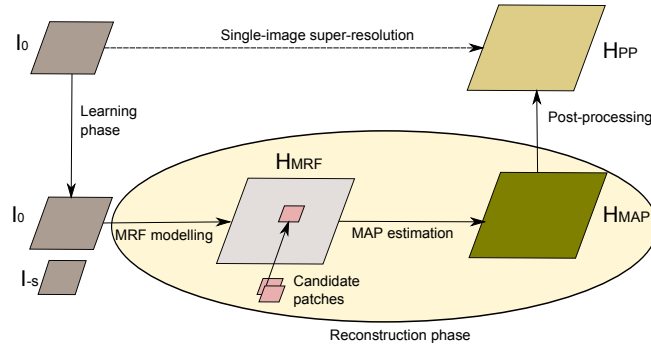
**Fig. 1.** The proposed single-image example-based super-resolution method.

message passing. Our results of the complete algorithm on different test images demonstrate a comparable or better performance than state-of-the-art SR techniques.

This paper is organized as follows. Sec. 2 describes our method for single image example-based SR, where Sec. 2.1 explains in more details the learning phase and Sec. 2.2 the reconstruction phase. Finally, we present and discuss the experimental results in Sec. 3 and we give our conclusion in Sec. 4.

2   Proposed Single Image Example-Based Method

We propose a single-image example-based super-resolution method which uses MRF to model the HR image as a collection of overlapping HR patches whose possible candidates are obtained from the input LR image itself. The algorithm can be divided into three main phases: learning, reconstruction and post-processing (see Fig. 1). In the *learning* phase, we find candidate patches of each unknown HR patch by first searching for k nearest neighbours of its corresponding known LR patch from the input image. This search exploits the patch redundancy across different scales of the Gaussian pyramid. We then extract the HR pairs of the found neighbours (called "parent" patches) from the input image and we use them as candidate patches for corresponding locations in the HR image, because we assume that the LR and HR patches are related in the same way across different scales. What follows is the *reconstruction* phase, which models the HR image as a MRF and performs inference on this model. MRF model has a great advantage over the simpler alternative, i.e. choosing the best match at each location, as we will demonstrate shortly.

Finally, we apply post-processing techniques to eliminate remaining artefacts. We use back-projection [1] to ensure the consistency of the HR result with the input LR image. In case of a small input image and high magnification factor, the search space may become too small for good matches to be found. This will result in visible artefacts so we also use steering kernel regression [11] that produces a smooth and artefact-free image while still preserving edges, ridges
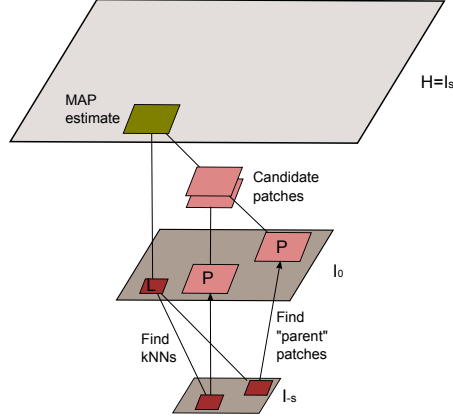
Fig. 2. An illustration of the process of learning candidate patches.

and blobs. Post-processing together with MRF modelling allows us to obtain competitive SR result even with only having LR image as the algorithm's input.

In the remaining of this section we will describe in details the learning and reconstruction phase.

2.1   Learning Candidate Patches

In this section we will explain how to use the single input image to obtain candidate patches. We use the example-based part of the algorithm from [8] in the sense that we search for similar patches within the Gaussian pyramid and use their "parent" HR patches for further reconstruction. However, our approach differs in the reconstruction step which enables us to perform simplified and faster search. Specifically, we search in only one level of the Gaussian pyramid whose sub-sampling factor is equal to the magnification factor for the reasons that will be explained shortly.

We start from the LR input image I_0 which is then blurred and sub-sampled with the integer factor s to yield the lower level of the Gaussian pyramid I_{-s}. The final goal is to reconstruct the image with a resolution that is s times higher than the original resolution. We will denote this HR image with $H = I_s$. The image I_{-s} will serve as a search space for matches of each patch from the image I_0. In details, the search and matching process has the following course, as illustrated in Fig. 2. For each pixel $p \in I_0$, where $p = (x, y)$ actually represents coordinates of the pixel in the image grid, we take its surrounding patch o_p (denoted by L on Fig. 2) and search for its k nearest neighbours (kNNs) in the image I_{-s}. Those neighbours are the patches that have the lowest sum of squared differences with o_p. Once kNNs y_p^n, $n = 1, .., k$, are found, their "parent" patches x_p^n, $n = 1, .., k$, (denoted by P on Fig. 2) are extracted from the given image I_0.

The "parent" patch represents a HR component of the HR-LR pair, where LR component is the LR patch. If the location of the central pixel of the LR
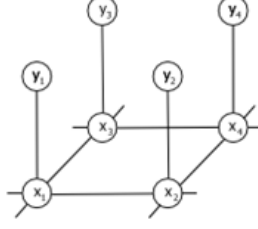
Fig. 3. MRF model: x_p are unknown HR patches and y_p measured LR patches.

patch y_p^n (found kNN) is $\tilde{p} \in I_{-s}$, then the location of the central pixel of the "parent" patch x_p^n is $s\tilde{p} \in I_0$ and its size is s times the size of the LR patch. These parent patches can now serve as candidate patches for each location sp in the HR image H which corresponds to the starting location p in the input LR image I_0. This is the reason for the same value s of the magnification and sub-sampling factor.

2.2   High-Resolution Image Reconstruction

After the algorithm described in the previous subsection, we have k candidate patches x_p^n, $n = 1, .., k$, for each location $sp \in H$. These locations correspond to starting locations $p \in I_0$ so we will refer to them with the index p. They are s pixels apart from each other in each direction in the HR grid. The naive approach would be to choose the best match, i.e. the nearest neighbour, at each location. Since the neighbouring patches will normally overlap, we can simply take the average in the overlap region. Although this solution could speed up the search process (because we only search for one nearest neighbour), the resulting image will have visible artefacts (Fig. 4).

Instead of just choosing the HR patch based on its agreement with the available data (the input image), we can take into account the relationship that inevitably exists between neighbouring locations in H in the sense that neighbouring patches should agree in the overlap region. This means that the sum of squared differences in the overlap region is minimal. Furthermore, we would like to observe the image as a whole rather than a collection of local assumptions. In this respect, we can formulate the choice of patches as a global optimization problem over the whole HR image by using the MRF framework [12]. For this purpose, we adopt the concept of [3] with a few major differences. First of all, our candidate patches are obtained from the input image itself, without using an external database. Moreover, they consist of raw pixel values instead of high frequency details so there is no need for preprocessing of the search space. Finally, we use our inference method for optimization which is simpler and faster than loopy belief propagation (LBP) [13] which was originally used.

Specifically, we model H as an undirected graph (Fig. 3) whose hidden nodes, indexed by p, represent the overlapping HR patches in the HR image that can

Fig. 4. Cropped version of zebra image 2x magnification. From left to right: best match result, MRF result with LBP as inference method, MRF result with NCMP as inference method.

take one of the values from the set $\{x_p\}$. Each hidden node is connected to the observed node (measured data) which is the LR patch o_p around pixel $p \in I_0$.

To completely define MRF model we still have to define compatibility functions between observed and hidden nodes (so called local evidence) and neighbouring hidden nodes. The former determines how much the unknown data agrees with measured data and the latter encodes prior information on the distribution of the unknown image. Local evidence is taken to be the Gaussian function of the matching error, i.e. sum of squared differences, between starting LR patch o_p and found k nearest neighbours y_p^n:

$$\phi_p(y_p^n, o_p) = \exp(-\|y_p^n - o_p\|^2/2\sigma_R^2), \tag{1}$$

Compatibility between neighbouring hidden nodes is the Gaussian function of the matching error in the region of overlap ROV of two neighbouring HR patches:

$$\psi_{p,q}(x_p^n, x_q^m) = \exp(-\|\mathrm{ROV}_{q,p}^n - \mathrm{ROV}_{p,q}^m\|^2/2\sigma_N^2). \tag{2}$$

σ_R and σ_N are the noise covariances which represent the difference between some "ideal" training samples and our image and training samples, respectively. Now, we have to choose one patch from the candidate set at each node that best fits the above constraints over the whole graph. This can be achieved by finding maximum a posteriori (MAP) estimates:

$$\hat{\mathbf{H}} = \hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|I_0) \tag{3}$$

$$P(\mathbf{x}|I_0) \propto \prod_{p,q} \psi_{p,q}(x_p^n, x_q^m) \prod_p \phi_p(y_p^n, o_p), \tag{4}$$

where ϕ_p is defined in equation 1 and $\psi_{p,q}$ in equation 2. This is generally a difficult problem to be solved exactly, but there is a number of approximate inference algorithms that can yield an approximate solution. We use our inference method called neighbourhood-consensus message passing (NCMP) [10] which is

simpler and and faster than LBP while the results are qualitatively very similar. Comparison of different approaches for HR image reconstruction is shown in Fig. 4. On the left we see the result of the best match approach which has a lot of artefacts due to its greedy nature. Using a MRF model produces much better result even if we use a simple inference method like NCMP (right image).

Fig. 5. Cropped castle image 2x magnification. From left to right and top to bottom: bi-cubic interpolation, MRF result, MRF with back-projection, MRF with back-projection and kernel regression.

3   Experimental Results

We tested our method on several images and compared it to the standard interpolation technique, like bi-cubic interpolation, and state-of-the-art SR techniques from [8], which is another single-image SR method, and [14], which uses a parametric learned edge model. In all experiments the LR patch size was 3x3 and HR patch size $3s$x$3s$, while parameters of MRF compatibility functions σ_R

Fig. 6. Cropped version of man image 2x magnification. From left to right: bi-cubic result, result of [14], result of the proposed method.

and σ_N slightly varied over different images. The number of nearest neighbours was $k = 10$.

In the first experiment, we demonstrate the effectiveness of our technique for sufficiently large search space. Fig. 5 shows the castle image and the results of our super-resolution algorithm with the magnification $s = 2$. It can be seen that the output of the MRF, without any post-processing, gives already reasonably good results. For example, all edges are sharp without "jaggy" artefacts which are visible in the result of bi-cubic interpolation. Back-projection further improves the result by eliminating artefacts and enhancing textures (e.g. texture on the roof). Finally, kernel regression only slightly smooths the image and can even be left out as a post-processing step in this case.

Table 1. RMSE and SSIM comparison of our method and bi-cubic interpolation result.

	norm. RMSE		SSIM	
Image	Our	Bi-cubic	Our	Bi-cubic
Zebras	0.3589	0.3948	0.9097	0.9043
Skyscraper	0.2573	0.2789	0.9275	0.9163
Butterfly	0.1371	0.1484	0.9572	0.9564

We also compare our result with two state-of-the-art methods from [8] and [14]. In Fig. 6 we can see that the proposed method eliminates "jaggies" along the lines present in the results of reference methods, e.g. lines on the collar of the sweater. Our method also outperforms reference methods for higher magnification factor, as shown in Fig. 7. It manages to produce the sharpest lines without "jaggy" or "ghosting" artefacts that are present in the results of [8] and [14], while keeping the result visually pleasing. Both results were obtained with
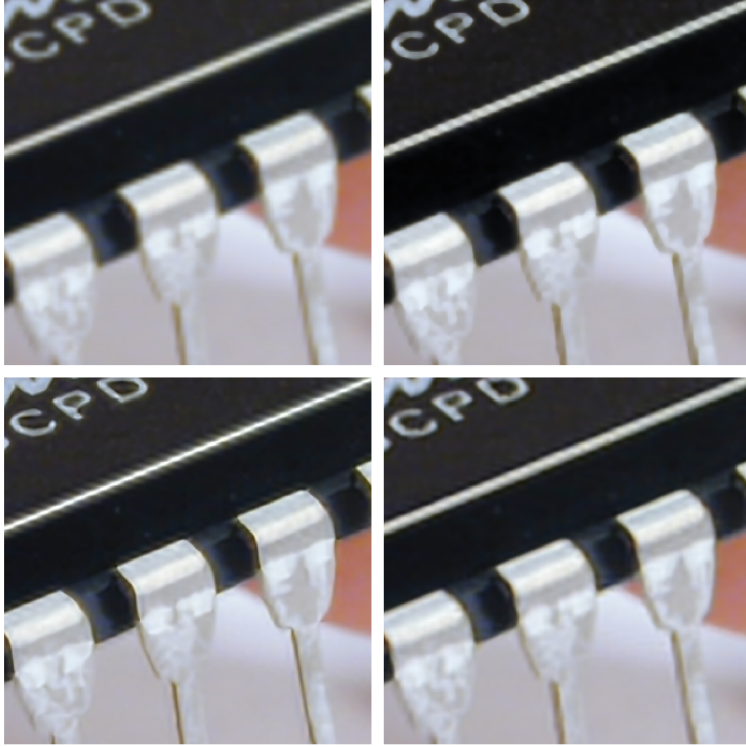
Fig. 7. Cropped version of chip image 4x magnification. Top-left: bi-cubic result. Top-right: result of [14]. Bottom-left: result of [8]. Bottom-right: result of the proposed method.

the input LR image of the small size. We believe that the difference would be even more significant for bigger input images.

In Table 1 we give quantitative results of a few images from Berkeley segmentation database [1]. We calculated the root mean square error (RMSE) and structure similarity index (SSIM) [15] between our super-resolution/bi-cubic interpolation result and ground truth. Our method produces smaller error and higher structure similarity score than bi-cubic interpolation. The quantitative improvement is, however, limited since the improvement is concentrated in edge regions, which represent small portion of the whole image.

4 Conclusion

In this paper we have presented a novel single-image super-resolution method based on MRF modelling. Unknown high-resolution image is modelled as a MRF

---

[1] eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench

whose nodes are overlapping high-resolution patches. Possible candidates for these nodes are found within only one level of the Gaussian pyramid of the input low-resolution image. To choose the best candidate in maximum a posteriori sense, we used our previously developed inference method called neighbourhood-consensus message passing, which makes this step fast and simple. Additionally, we performed back-projection and steering kernel regression to further improve the results. Results show that our method greatly outperforms standard techniques, while being visually better or comparable with state-of-the-art techniques.

References

1. Irani, M., Peleg, S.: Improving Resolution by Image Registration. CVGIP: Graphical Model and Image Processing, vol. 53, no. 3, 231–239 (1991)
2. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and Robust Multi-Frame Super-Resolution. IEEE Trans. Image Proc., vol. 13, no. 10, 1327–1344 (2004)
3. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning Low-Level Vision. IJCV, vol. 40, no. 1, 24–47 (2000)
4. Baker, S., Kanade, T.: Limits on Super-Resolution and How to Break Them. PAMI, vol. 24, no. 9, 1167–1183 (2002)
5. Tappen, M.F., Freeman, W.T.: Exploiting the Sparse Derivative Prior for Super-Resolution and Image Demosaicing. In: IEEE Workshop on Stat. and Comp. Theories of Vision (2003)
6. Kim, K., Kwon, Y.: Example-Based Learning for Single-Image Super-Resolution. In: Proc. of the 30th DAGM symposium on Pat. Rec. LNCS, vol. 5096, pp. 456–465. Springer, Heidelberg (2008)
7. Ebrahimi, M., Vrscay, E.: Solving the Inverse Problem of Image Zooming Using ”Self-Examples”. In: Proc. of Int. Conf. on Image Analysis and Recognition (ICIAR). LNCS, vol. 4633, pp. 117–130. Springer-Verlag (2007)
8. Glasner, D., Bagon, S., Irani, M.: Super-Resolution from a Single Image. In: Int. Conf. on Comp. Vision (ICCV) (2009)
9. Luong, H., Ružić T., Pižurica, A., Philips, W.: Single Image Super-Resolution Using Sparsity Constraints and Non-Local Similarities at Multiple Resolution Scales. In: SPIE (2010)
10. Ružić T., Pižurica, A., Philips, W.: Neighbourhood-Consensus Message Passing and Its Potentials in Image Processing Applications. In: SPIE Electronic Imaging (2011)
11. Takeda, H., Farsiu, S., Milanfar, P.: Kernel Regression for Image Processing and Reconstruction. IEEE Trans. Image Proc., vol. 16, 349–366 (2007)
12. Li, S.Z.: Markov Random Field Modeling in Computer Vision. Springer-Verlag (1995)
13. Yedidia, J.S., Freeman, W.T.: On the Optimality of Solutions of the Max-Product Belief-Propagation Algorithm in Arbitrary Graphs. IEEE Trans. Inf. Theory, vol. 47, no. 2, 736–744 (2001)
14. Fattal, R.: Image Upsampling via Imposed Edge Statistics. In: SIGGRAPH ’07 (2007)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Proc., vol. 13, no. 4, 600–612 (2004)