

時系列低照度画像による高輝度画像の復元

林正紘[†] 坂上文彦[†] 佐藤淳[†] 是枝義輝[‡] 東久保政勝[‡] 山本英典[‡]

名古屋工業大学[†]

住友電気システムソリューション株式会社[‡]

1. はじめに

夜間などの低照度環境では、自動車のように移動する物体を鮮明に撮影することは困難である。十分な光量を得るために長時間露光で撮影した場合には被写体の動きによるぶれが発生し、ぶれを避けるために短時間露光で撮影した場合には大きな画像ノイズが発生する。このようなぶれやノイズが発生すると、自動車において特に重要な情報であるナンバープレートを読み取ることが困難となる。このような問題を解決するため、特殊な装置を用いた撮影技術や、劣化画像から高画質な画像を復元する画像処理技術が提案されている。しかし、特殊な装置を使用する方法は高価であり、使用用途が限定される。そこで本研究では、深層学習を用いて、劣化した時系列低照度画像から、鮮明な高輝度画像を復元する手法を提案する。時系列画像を用いることで、被写体に関する独立した情報をより多く得ることができ、より鮮明な高輝度画像の復元が可能となる。また、文字識別損失という新しい損失関数を導入して学習を行うことで、復元した高輝度画像中の文字の可読性をより向上させる。

2. 提案手法

提案手法の概要を図1に示す。この図が示すように、提案手法では運動物体を復元するU-Net、静止物体を復元するU-Net、運動物体のマスク画像を生成するU-Netの3のネットを用いる。

2.1 運動物体の位置合わせ

時系列画像を扱う場合、各時刻間における対応点の移動は重要な情報である。時刻1から時刻 T までの T 時刻の低照度画像 I_i ($i = 1, \dots, T$)から時刻 T における高輝度画像を復元する場合、対応点のオプティカルフローが既知であれば、より高精度な復元ができると考えられる。そこ

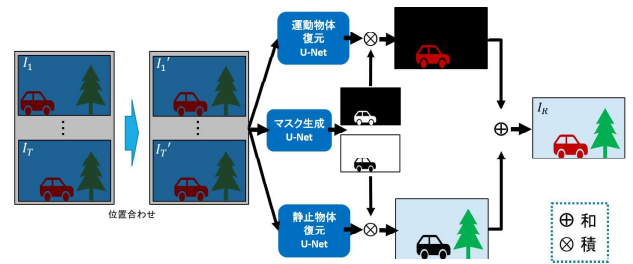


図1：提案手法のネットワークの概要

で本研究では、時系列画像中の運動物体に関するオプティカルフローを事前に推定しておき、それをもとに大まかな位置合わせを行う。

2.2 時系列画像による画像の復元

画像ノイズは一般に撮影ごとにランダムに発生するものなので、同じ環境下で撮影した場合であっても、撮影のたびに異なるノイズが乗った画像が得られる。このため、多時刻の低照度画像を用いることで、単一の低照度画像のみを用いる場合よりも被写体の情報をより多く得ることができる。本研究では、時刻1から時刻 T までの T 時刻の低照度画像 I_i ($i = 1, \dots, T$)を入力とし、最終フレームである時刻 T における高輝度な復元画像 I_R を出力とするネットワークを考える。本研究では、このようなネットワークとしてU-Netを用いる。U-Netの入力は、時系列低照度画像をチャンネル方向へ連結させたもの、出力は復元された高輝度画像とする。画像の縦の画素数を H 、横の画素数を W 、チャンネル数を C 、時刻数を T とすると、ネットワークへの入力大きさは $H \times W \times C \times T$ となる。また、出力は $H \times W \times C$ である。

2.3 複数ネットワークによる画像の復元

時系列画像を用いた復元では、一般に運動物体の復元は静止物体の復元より難しい。これは、静止物体が画像間で同じ位置で観測されるのに対して、運動物体は位置合わせをしてもわずかに異なる位置で観測されるためである。そこで本研究では、運動物体の復元を学習するU-Netと静止物体の復元を学習するU-Netの2つのネットワークを構築し、画像全体の復元をより高精度に行うことを考える。さらにこれら2つのU-Net

への入力を制御するために、運動物体を抽出するマスク画像と静止物体を抽出するマスク画像を生成するU-Netを導入する。運動物体復元U-Netと静止物体復元U-Netの出力画像に対して、対応するマスク画像によるマスク処理を行うことで、運動物体と静止物体の復元を別々に学習することができる。静止物体復元U-Netの損失関数は正解画像と出力画像のL1損失とし、運動物体復元U-Netの損失関数は、L1損失に次の章で述べる文字識別損失を加えたものとする。また、マスク生成U-Netの損失関数は正解画像と出力画像から計算されるBinary Cross Entropy損失とする。

2.4 文字識別損失

一般的な損失関数であるMAEやMSEでは、文字の骨格情報を評価することが難しい。そこで本研究では、車両において重要なナンバープレートの復元を効率的に学習するために、新たに文字識別損失を導入する。文字識別損失は、学習済みの文字識別器へ正解画像と復元画像を入力し、それぞれの識別結果から計算する。本研究では、物体認識器Retina Net[1]をThe Street View House Numbers Dataset[2]で学習済みの識別器を使用する。

3. 実験

提案法の有効性を確認するための実験結果を示す。本研究では、高輝度正解画像から仮想的に画像ノイズの乗った低照度画像を生成することで、低照度画像と高輝度画像のペアよりなるデータセットを構築した。高輝度正解画像のデータは、固定カメラを用いて走行する車両を撮影した連続画像データを530車種分用意した。その内、500車種を訓練データ、30車種をテストデータとし、各データに対して5通りの仮想低照度画像を生成しネットワークの学習を行った。

図2にテストデータの入力画像、図3に復元結果を示す。復元画像のナンバープレートに注目すると、単一画像による復元よりも、時系列画像を用いることで復元精度が向上し、文字識別損失を加えることでさらに精度が向上していることがわかる。

次に、提案法の復元精度に関して定量評価を行った結果を示す。本研究ではテストデータのナンバープレートの復元精度に着目し、復元画像のナンバープレートの文字が学習済みの文字識別器YOLOv3[3]によってどれだけ正しく識別されるかを評価した。表1にテストデータの正答率を求めた結果を示す。この表より、時系列画像と文字識別損失を使用することで正答率が大幅に向上しており、提案法の有効性が確認できる。

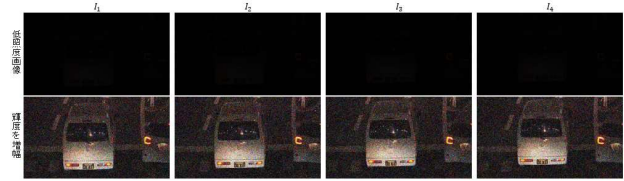


図2: 入力時系列画像. 上段は元の低照度画像. 下段は上段の画像の輝度を単に増幅したもの.



図3: 復元結果. (左上) 正解画像 (右上) 単一画像による復元 (左下) 4枚の画像による復元 (右下) 文字識別損失を加えた4枚の画像による復元

表1: 訓練済みの文字識別器による正答率. 左側の数字は入力画像数を表す. また, L_1 は L1 損失, L_{RC} は文字識別損失を表す.

	正答率
1, L_1	0.591
2, L_1	0.774
4, L_1	0.816
1, $L_1 + L_{RC}$	0.725
2, $L_1 + L_{RC}$	0.826
4, $L_1 + L_{RC}$	0.880

4. まとめ

本稿では、時系列低照度画像から高輝度画像を復元する手法を提案し、その有効性を確認した。

参考文献

- [1] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P., Focal loss for dense object detection. Proc. IEEE CVPR, pp. 2980–2988, 2017.
- [2] Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A. Y., Reading digits in natural images with unsupervised feature learning. Proc. NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.
- [3] Redmon, J., & Farhadi, A., Yolov3: An incremental improvement. arXiv:1804.02767, 2018.

論文の利用上の注意

本論文の著作権は、情報処理学会に帰属します。本論文は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。