社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

画像処理による単純な文字の特徴の増加手法の提案

津山　裕加†　岩村　雅一†　黄瀬　浩一†

† 大阪府立大学大学院工学研究科
〒 599-8531 大阪府堺市中区学園町 1-1
E-mail: tsuyama@m.cs.osakafu-u.ac.jp, {masa,kise}@cs.osakafu-u.ac.jp

あらまし　近年，情景中の文字を認識する技術についての研究が盛んに行われている．既存の情景内文字認識手法の一つである松田らの手法では，画像から得られる局所特徴を用いることで，複雑背景・レイアウト上の文字でも高精度で認識することができる．しかし，数字やアルファベットといった単純な文字は，局所特徴が十分に得られないことから認識率が低い．そこで本稿では，認識対象の画像に画像処理を施し，文字の特徴を残したままぼかすことによって文字領域から有効な特徴をより多く抽出し，単純な文字の認識を可能にする手法を提案する．実験の結果，ぼかすことで新たに認識できる文字があることが分かった．

キーワード　情景内文字認識，局所特徴，符号化開口

1. はじめに

近年，カメラを搭載した高性能なモバイル型端末の普及に伴い，それらを用いて情景中の文字を認識する技術が注目されている．この技術により，ユーザがカメラで撮影することで得られた画像に写っている文字を自動で認識することができ，更にはその文字に関連する情報を検索・提供するといったサービスの実現が可能になる．これにより，ユーザは文字を自分で入力する必要がなくなる．対象となる文字の数が非常に多い場合や文字が複雑である場合，あるいはユーザ自身が分からない言語の場合でも，その文字を撮影するだけで単語や関連情報を検索することができる．

しかし，現在の文字認識技術では情景画像中の文字に対して認識可能な文字が限られているという問題がある．これは認識対象となる情景中文字が多様であることが原因である．情景画像中の文字は形状が多様であり，また，カメラで撮影した際の照明の強さや撮影対象との角度，画像中に映りこんだ影や物体，背景，あるいは画像の解像度などの条件によって画像が変化してしまうことも，情景画像中の文字認識を困難にしている．

情景画像中の文字を認識するためには文字の多様性に対応する必要がある．そのような情景内文字を認識する試みは数多くなされているが，そのうちのほとんどがラテン文字のみを対象としている．

それに対して松田らの手法 [1] は，ラテン文字に限らず漢字やひらがなのような日本語も認識対象とした情景内文字認識手法である．画像の部分的な形状を反映した局所特徴を用いることで，照明変化に強く，複雑な背景やレイアウトの文字であっても認識することができる．この松田らの手法は，漢字のような複雑な文字の認識率は高いが，数字やアルファベットといった単純な文字の認識率はあまり高くない．これは，文字の形が単純であるため文字領域から認識に有効な局所特徴があまり得られず，文字をうまく検出できなかったり，誤認識を起こしてしまうからである．具体的には，松田らの手法であれば一直線上にない三点が認識に必要な特徴点数であるが，単純な文字からはこれらを得ることが難しく，認識がうまくいかないことがある．

そこで本研究では，単純な文字から認識に有効な特徴をより多く抽出することを目的とする．これを実現するため，単純な文字から得られる特徴量を増やす方法を検討したところ，伊村らの手法 [2] が参考になるのではないかと考えた．伊村らの手法では，コンピュテーショナルフォトグラフィの技術のひとつである符号化開口をパターン認識に特化した条件で用いることで，単眼カメラで撮影された，ぼけた文字の認識を可能にする．この手法を用いた実験において，符号化開口を用いてぼかした画像からはぼかす前の画像よりも多くの特徴が得られたと報告されている．伊村らの手法はカメラでパターンを撮像する際に起こる光学的な現象を扱っていたが，任意の画像にこれと同じ処理を適用することで，得られる特徴数を増加させることができると考えられる．

そこで本稿では，符号化開口の一種である Veeraraghavan の画像 [3] を文字画像に畳み込み，文字の特徴を残したままぼかすことによって文字領域から有効な特徴をより多く抽出し，松田らの手法において単純な文字の認識を可能にする手法を提案する．本稿ではさらに，ぼかすことで認識にどのような影響を与えるかを調査する．具体的には以下の通りである．局所特徴を得る処理は特徴点決定と特徴量抽出の２つに分割できる．画像がぼけた場合，それらのどちらが認識に寄与するのかを調査する．

2. 関連研究

2.1 局所特徴

局所特徴とは，画像の一部から得られる特徴量のことであ

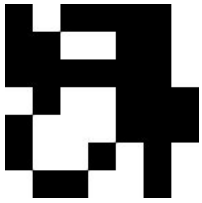


図 1　Veeraraghavan の開口形状

る．情景画像中の文字認識における課題は，カメラで撮影した際の照明の強さや撮影対象との角度，画像中に写りこんだ影や物体，背景，あるいは画像の解像度などの条件によって画像が変化してしまうことである．そこで，局所特徴を用いることで，画像の一部が変化していても残りの部分から得られた特徴量によって正しく識別できると考えられる．今日では様々な局所特徴が提案されている．そのような特徴の一つに Scale-Invariant Feature Transform（SIFT）[4] がある．SIFT とは，D.Lowe が考案した特徴量で，回転やスケール変化，照明変化に強いという利点がある．

局所特徴を得るためには，大きく分けて 2 つの手順を踏む必要がある．まず 1 つは，与えられた画像の中から特徴量を計算する領域を検出すること（特徴点決定）である．もう 1 つは，その領域ごとに特徴量を計算すること（特徴量抽出）である．前者の特徴点決定であるが，SIFT では Difference of Gaussian（DOG）を用いている．これにより，画像中の特徴的な部分を検出し，そこから特徴抽出を行っている．他の特徴点決定方法としては，Dense Sampling と呼ばれるものがある．Dense Sampling はあらかじめ特徴点の領域と数を決定しておく方法であり，どのような画像であっても一定数の特徴点を得ることができる．

2.2　伊村らの手法

本節では提案手法に関連した手法として伊村らの手法について述べる．一般的なレンズカメラでは，絞りの形が円形であることから，ぼけの形状も円となる．そのため，ピントがずれた時には，ぼけによって対象の特徴が潰れてしまい，ぼけを除去しても鮮明な画像を得ることができない．一方，ピンホールカメラを用いるとぼけは発生しないが，光量が不足して全体的に暗い画像となってしまい，画像中のノイズが強調されてしまう．符号化開口は光量を確保しつつ，ぼけの除去後に鮮明な画像を得るために設計された．符号化開口のひとつである Veeraraghavan の開口形状を図 1 に示す．符号化開口は高周波成分が残りやすいように設計されているため，劣化パターンにもパターンの持つ特徴が残りやすくなる．一般的に符号化開口を用いた認識手法では，ぼけを除去した鮮明な画像を作成してから，パターン認識できると考えられる．一方で伊村らの手法では，データベースには様々な程度のぼけを生じさせた画像を登録しておき，データベースに存在するテンプレートのいずれかが写っていると仮定することで，ぼけた画像の復元を行わずに認識を可能とする．

伊村らの実験では，符号化開口でぼかした文字画像からぼかす前よりも多くの局所特徴が得られるという知見が得られた．この実験では撮像時に生じるぼけをシミュレーションしていたが，与えられた任意のクエリ画像を同様の処理でぼかすことで，画像から得られる局所特徴を変化させることができると考えられる．そこで提案手法では，図 1 の Veeraraghavan の画像を用いて文字画像に畳み込みを行う．

3.　松田らの手法 [1]

本節では松田らの手法について述べる．松田らの手法は，局所特徴を利用した認識手法である．局所特徴にスケールと回転に不変な SIFT 特徴を使っており，加えて特徴点の対応を取った後にアフィン変換行列を計算することで，スケール変化や回転に頑健な文字認識を実現している．文字単位の認識が可能であるため，直線上に配置されていない，複雑なレイアウト上にある文字であっても認識ができる．また，局所特徴は画像の局所領域から得られることから，背景による影響が少ないため，複雑な背景上にある文字でも認識できる．

松田らの手法は，認識を行う前にデータベースの構築と学習を行なっている．データベースの構築では，データベース用の文字画像から SIFT 特徴を抽出し，データベースに登録する．松田らの手法では，抽出した特徴にどの文字から抽出されたかという情報だけでなく，文字の中心との位置関係（Reference Point:RP）も保持している．認識する場合も同様に，クエリから SIFT 特徴を抽出し，抽出された特徴ごとにデータベースに登録されている特徴とのマッチングを行う．特徴ごとにマッチングを行った結果，クエリの特徴ごとにデータベース内の特徴と関係づけられ，クエリ上に各特徴の RP が射影される．そして射影された RP が一定領域内に閾値以上存在した時，それを文字とみなして領域を確定し，認識結果を返す．このように，画像中の文字領域の事前の切り出しを必要とせずに認識できるのも松田らの手法の特徴である．

4.　提 案 手 法

松田らの手法は，漢字のように複雑な文字に対しては高い識別性能を示す．これは，複雑な文字は形状が複雑であることから認識に必要な局所特徴を十分に得られるためである．しかし，数字やアルファベットのように単純な文字の場合，形状の単純さゆえ局所特徴があまり得られない．そのため，単純な文字に対しては認識率が低くなるという問題がある．そこで，単純な文字から得られる有効な特徴を増やすことが必要となる．本稿では，符号化開口の一種である Veeraraghavan の画像 [3] を用いてぼかし，元の文字の特徴を残しつつ，得られる特徴の数を増やす手法を提案する．実際に畳み込んだ結果を図 3 に示す．図 3 を見ると，元の文字（この場合は数字の 7）の特徴を残しながらぼけていることがわかる．これにより，その文字を認識するのに有効な特徴を増加させることができると考えられる．

提案手法を用いた認識システムの概要を図 2 に示す．図 2 のように，あるクエリが入力として与えられた時，そのクエリを複数の識別器で認識する．具体的には，ぼかさずにそのまま認識するものと，ぼかして認識するものとがあり，ぼかす場合はぼけの大きさに準じて識別器の数が増加する．複数の識別器で

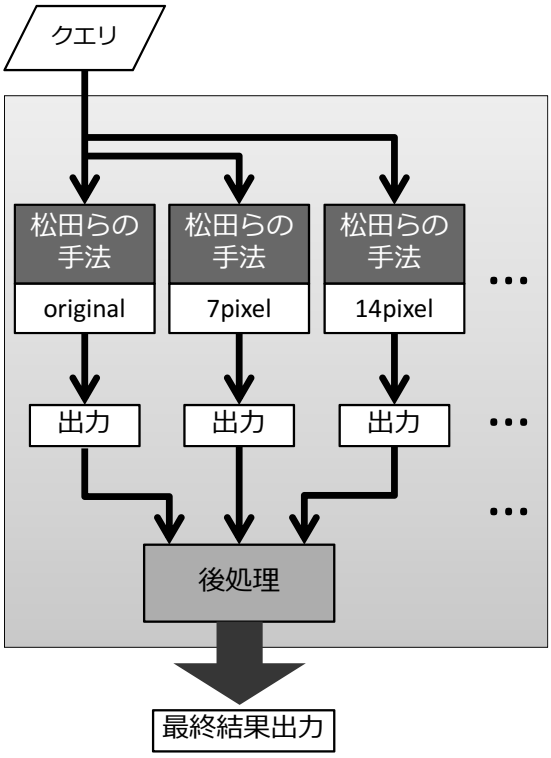


図 2　提案手法を用いた認識システムの概要

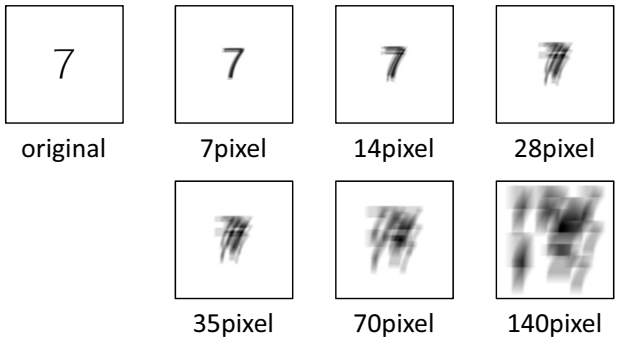


図 3　畳み込んだ例

それぞれ認識結果が得られるので，その中でよいものを採用し，結果を統合する．その統合した結果を最終的な認識結果として返すのが，提案手法の枠組みである．ただし，後処理については具体的にどのように行うかは今のところ検討段階である．そのため，本稿ではいずれかの識別器で正解が得られたら正解とすることで提案手法の可能性の検証を行うのみにとどめる．

提案手法では，ぼけた画像に対して SIFT の特徴点検出・特徴抽出を行う．この時，ぼけた画像は特徴点検出と特徴抽出の一方のみ，あるいは両方に使用する．つまり，ぼけた画像から得られた特徴点に基づいて元の画像から特徴を抽出するものと，ぼけた画像から得られた特徴点に基づいて元の画像から特徴を抽出するものと，ぼけた画像から得られた特徴点に基づいてぼけた画像から特徴を抽出するものの 3 パターンについて実験を行うことで，ぼかすことが特徴点決定・特徴抽出のどちらに影響があるのかを調査する．

松田らの手法は本来，認識対象のアフィン歪みに頑健である．しかし，今回用いた Veeraraghavan の開口形状が回転不変でないために，それと畳み込んでぼかしたパターンの回転に対する頑健性が失われることが知られている [2]．この問題に対処するために，回転不変な符号化開口を作成し，それを用いて実験することは今後の課題である．

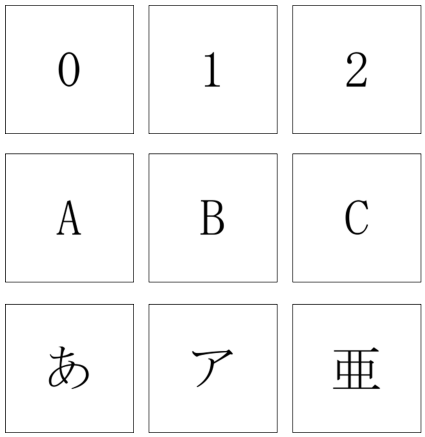


図 4　学習に使用した文字画像の例

5.　実験・考察

本節では実験により提案手法の性能を調査する．

5.1　実 験 条 件

学習用画像として，MS 明朝の数字，アルファベット，ひらがな，カタカナ，漢字の計 3968 文字の画像を用いた．画像の大きさはすべて 1 辺 200[pixel] で，文字の大きさは縦が約 55[pixel] である．またいずれも二値画像である．図 4 に一例を示す．また，畳み込みに用いる Veeraraghavan の画像は，1 辺の長さがそれぞれ 7, 14, 28, 35, 70, 140[pixel] の計 6 枚を用意した．

また，クエリ画像として，以下の手順で用意した撮影画像を用いた．文字の座標，回転角度，大きさが既知になるように配置した電子画像を作成し，作成した電子画像を印刷し，それをカメラで撮影した．電子画像の背景には，本稿では無地の背景を用意した．そしてその背景に対して，MS 明朝のフォントを用いて，英数字とひらがな，カタカナ，漢字の 3968 文字を 52 枚に分割して描画した．同時に，電子画像の周りに 5 つのマーカーとページ番号と QR コードを描画した．5 つのマーカーは，撮影された画像上における電子画像の位置を特定するために用いた．また QR コードは，撮影している画像が何枚目の電子画像に対応するかを自動的に取得するために用いた．これらはいずれも正解データを自動的に生成するためだけに用い，認識時には 利用しないので，撮影後のクエリ画像ではこれらを白で塗りつぶしてある．これら電子画像数は合計 2808 枚に上り，これらを撮影するためには非常に時間がかかるため，環境照明の変化による影響が懸念された．そこで，直接照明を近距離から当てることで，照明変化による影響を最小限に留めるようにした．また，撮影に用いたカメラは，正面に配置した．これにより得られたクエリ画像は合計 2808 枚であった．図 5 はクエリ画像の例である．クエリ画像の大きさは全て縦が 1080[pixel], 横が 1920[pixel] となっている．

文字の大きさであるが，縦が約 60[pixel] で，学習用画像の文

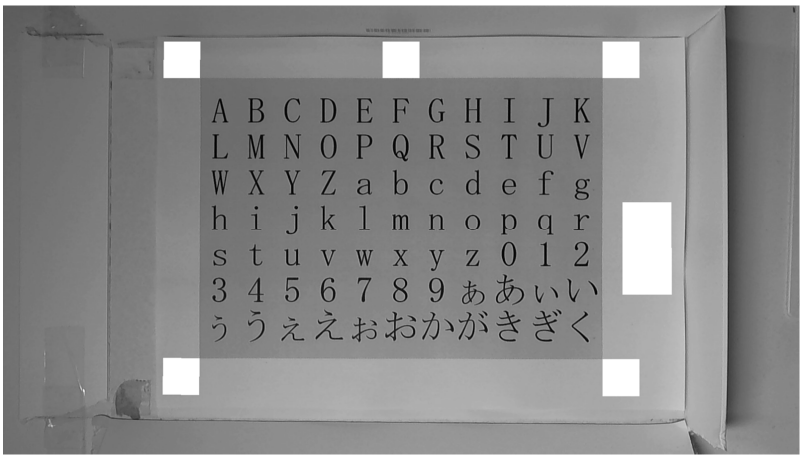


図 5　クエリ画像の例

表 2　実験に用いた手法と使用した画像

	detector	descriptor
既存手法（SIFT)	original	original
既存手法（DenseSIFT）	original	original
提案手法 1	original	ぼかした画像
提案手法 2	ぼかした画像	original
提案手法 3	ぼかした画像	ぼかした画像

字の約 1.1 倍の大きさとなっている．文字の大きさが異なっていると，主に曲線部分から得られる特徴は変化してしまう．それにより，曲線的な文字の認識率に影響があると考えられる．そのため，同じ文字の大きさでの実験を今後の課題とする．

5.2　実　験　1

実験 1 では，畳み込みによって文字画像から得られる特徴点数がどのように変化するかを調査するため，計 3968 枚の学習用画像に対して畳み込みを行い，特徴点数の集計を行った．畳み込みについては，OpenCV の関数を使用した．特徴点検出手法 (detector)，特徴抽出手法（descriptor），ならびにそれぞれに用いる画像は表 1 の通りである．

Dense Sampling のパラメータは，スケールを 50，点の間隔を 20 と設定した．他にもいくつかのパラメータで試したが，認識結果にあまり差が出なかったことから本稿ではこのパラメータを採用した．以降，実験に用いた手法を，表 2 のように呼ぶ．

実験結果を表 3 に示す．提案手法 2 と提案手法 3 は特徴点検出に用いる画像が同じであるため，同じ結果となる．表 3 より，畳み込みに用いる画像の大きさが 7[pixel] のときと 14[pixel] のときは，提案手法 2 および提案手法 3 において全体の特徴点数が増加することが分かった．28[pixel] 以上の大きさの画像で畳み込んだ場合は特徴点数が減少しているが，これはぼけが大きくなるにつれて，主に漢字から得られる特徴点数が大幅に減少したためと考えられる．漢字はもともと複雑な形状をしているため，ぼけが少し大きくなるだけで隣の文字や自分自身との干渉が大きくなり，特徴が得られないように変化してしまったのだと考えられる．今後の課題として，クエリの文字の大きさに応じて適切なぼけの大きさを調査する必要がある．また，どの程度の間隔があれば干渉を無視できるかというのも調べる必要がある．

表 4　松田らの手法のパラメータ

パラメータの内容	パラメータの値
対応点探索の探索数	15
認識に必要な最小点数	4
RP のクラスタリングの半径	16

5.3　実験 2：得られた特徴の性能評価

実験 2 では，抽出した特徴が認識に有効であるかを確認するため，松田らの手法で認識した時の認識率を調査した．認識性能の評価には，適合率と再現率，F 値を使用する．F 値は，適合率と再現率の総合的な評価の際に使用される評価基準であり，以下の式で表される．

$$F\ 値 = \frac{2 \cdot 再現率 \cdot 適合率}{再現率 + 適合率} \tag{1}$$

松田らの手法のパラメータは表 4 のように設定した．

結果を表 5.3〜表 5.3 に示す．表 5.3 は再現率 [%]，表 5.3 は適合率 [%]，表 5.3 は F 値をそれぞれ示している．表 5.3 から，提案手法 1 で 14[pixel] の画像を使った時，提案手法 2 で 7[pixel] の画像を使った時，そして提案手法 3 で 7[pixel] の画像を使った時，既存手法よりも識別性能が高くなると分かった．加えて，畳み込みに用いる画像が 28[pixel] 以上になると，識別性能は低下していることが分かる．これは，認識対象の文字の大きさ，および文字同士の間隔に対してぼけが大きすぎるため，干渉が大きくなり文字の特徴が認識に適さなくなっただと考えられる．また，既存手法（Dense SIFT）では文字をほとんど検出，認識できなかった．図 6 は認識結果の一例である．既存手法（SIFT）では認識できていなかった一部の単純な文字が，新たに検出，あるいは認識できていることがわかる．また，既存手法 (Dense SIFT) では特徴点数が多く検出した文字数も多いが，誤検出がかなり多いことも図よりわかる．

6.　まとめと今後の課題

局所特徴を用いて文字認識する際の課題である，単純な文字から得られる特徴が少ないという問題に対して，符号化開口をヒントに画像に処理を加えてから特徴抽出を行うことで対処を

表 1　実験で用いた特徴点検出手法，特徴抽出手法，および使用画像

特徴点検出手法（detector）	SIFT	Dense Sampling(既存手法のみ)
特徴抽出手法（descriptor）	SIFT	
使用する画像	ぼかしていない画像（original）	ぼかした画像 (7，14，28，35，70，140[pixel])

表 3　実験 1 の結果：得られた特徴点数

	使用した画像						
	original	7[pixel]	14[pixel]	28[pixel]	35[pixel]	70[pixel]	140[pixel]
既存手法（SIFT）	248707	-	-	-	-	-	-
既存手法（Dense SIFT）	394200	-	-	-	-	-	-
提案手法 1	-	248707	248707	248707	248707	248707	248707
提案手法 2 ・提案手法 3	-	250850	452162	344859	297308	143629	80281

表 5　再現率 [%]

	使用した画像						
	original	7[pixel]	14[pixel]	28[pixel]	35[pixel]	70[pixel]	140[pixel]
既存手法（SIFT）	87.25	-	-	-	-	-	-
既存手法（DenseSIFT）	0.08	-	-	-	-	-	-
提案手法 1	-	87.32	89.84	86.24	80.19	8.95	0.00
提案手法 2	-	90.85	69.98	8.69	0.05	0.00	0.00
提案手法 3	-	90.65	68.83	4.31	0.00	0.00	0.00

表 6　適合率 [%]

	使用した画像						
	original	7[pixel]	14[pixel]	28[pixel]	35[pixel]	70[pixel]	140[pixel]
既存手法（SIFT）	94.93	-	-	-	-	-	-
既存手法（DenseSIFT）	0.04	-	-	-	-	-	-
提案手法 1	-	93.98	95.70	95.27	92.58	73.65	0.00
提案手法 2	-	92.96	89.26	88.69	100.00	0.00	0.00
提案手法 3	-	93.19	89.69	85.93	0.00	0.00	0.00

表 7　F 値．1 に近ければ近いほど高性能である．

	使用した画像						
	original	7[pixel]	14[pixel]	28[pixel]	35[pixel]	70[pixel]	140[pixel]
既存手法（SIFT）	0.91	-	-	-	-	-	-
既存手法（DenseSIFT）	0.00	-	-	-	-	-	-
提案手法 1	-	0.91	0.93	0.91	0.86	0.16	0.00
提案手法 2	-	0.92	0.78	0.16	0.00	0.00	0.00
提案手法 3	-	0.92	0.78	0.08	0.00	0.00	0.00

試みた．具体的には，Veeraraghavan の画像を畳み込み文字の特徴を残したままぼかすことで，単純な文字から得られる，認識に有効な特徴を増やすことを試みた．

実験結果より，事前の画像処理によって特徴点数を増加させることは可能であることがわかった．また，実際に認識を行った結果，本稿で用いたクエリに対しては，7[pixel]，14[pixel] の大きさの画像でぼかすと認識精度を向上させることが出来ると分かった．

今後の課題として，まず他のフォントでも同様の実験をすることが挙げられる．また，今回は元画像にのみ Dense Sampling を用いたが，ぼけた画像に対して Dense Sampling をした場合についても調査する必要がある．加えて，文字の大きさやクエリの大きさ，あるいは文字の間隔などに応じてぼけの大きさを変化させることで，より高精度で認識することが可能であると考えられる．そのため，適切なパラメータを調査する必要がある．さらに，より複雑な背景上，あるいはレイアウトの文字を認識できるかどうかについての調査も必要である．

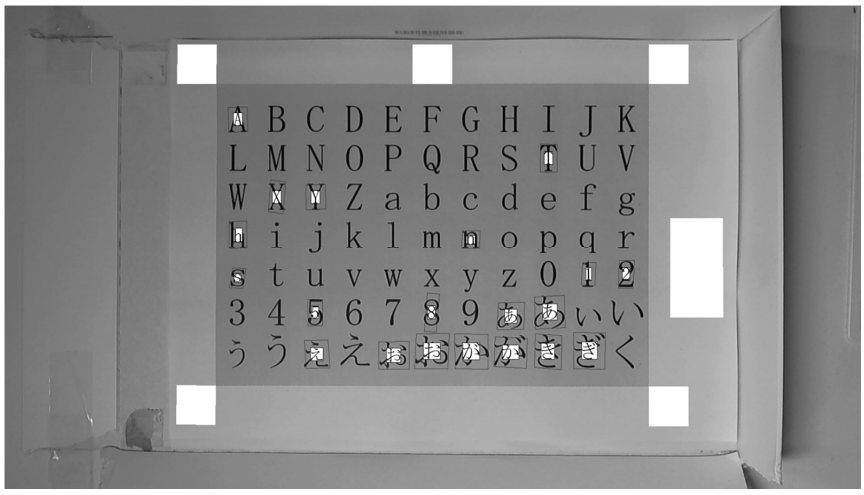
謝辞　本研究は， JST CREST の補助による．

文　献

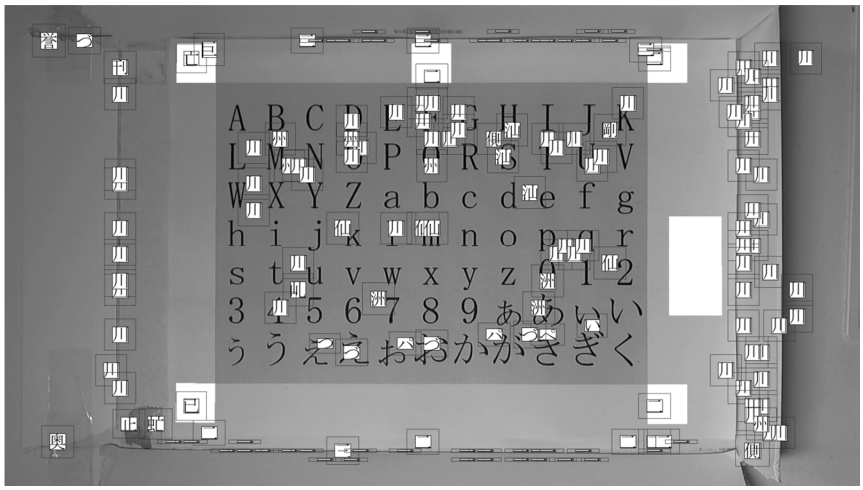
[1] T. Matsuda, M. Iwamura, and K. Kise, “Performance improvement in local feature based camera-captured character recognition,” Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS2014), pp.196–201, April 2014.

[2] M. Iwamura, M. Imura, S. Hiura, and K. Kise, “Recognition of defocused patterns,” IPSJ Transactions on Computer Vision and Applicaitons (CVA), vol.6, pp.48–52, July 2014.

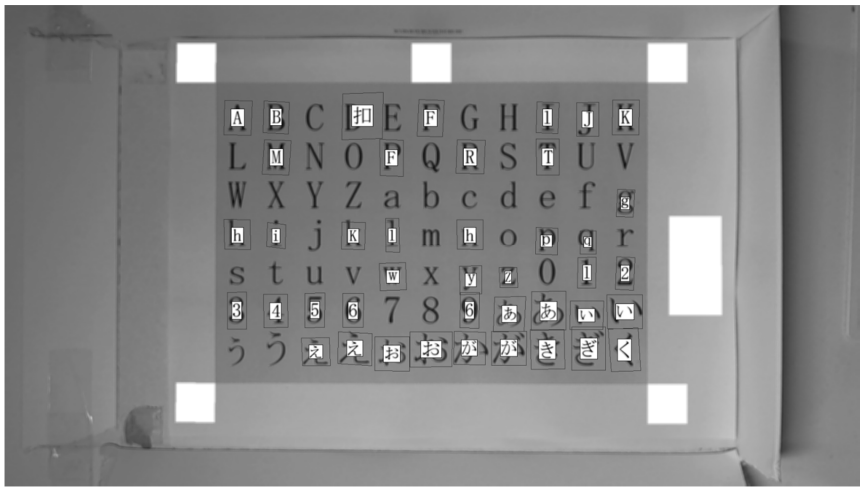
[3] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J.



(a) 既存手法（SIFT）



(b) 既存手法（Dense SIFT）



(c) 提案手法 3（7pixel のとき）

図 6　認識結果の例

Tumblin, “Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” ACM Trans. Graph., vol.26, no.3, p.69, 2007.

[4] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” IJCV, vol.60, no.2, pp.91–110, 2004.